

(Founded in 1935 by Carl Murchison)

# The Journal of PSYCHOLOGY

The General Field of Psychology

GORDON ALLPORT  
Harvard University

JOHN A. ANTOINETTI  
Hartford, Connecticut

ARTHUR G. BILLS  
University of Cincinnati

E. A. BOTT  
University of Toronto

LEONARD CARMICHAEL  
Smithsonian Institution

HERBERT S. CONRAD  
U. S. Office of Education

JOHN F. DASHIELL  
Florida State University

ROY M. DORCUS  
University of California  
at Los Angeles

FRANKLIN FEARING  
University of California  
at Los Angeles

FRANK A. GELDARD  
Princeton University

If this space should be unstamped, this is the regular library edition.  
But if this space is stamped with a designating title, this is a special

edition, sold under the restrictions of a bilateral contract, and may not  
be resold for a period of five years from date of publication.

CLARENCE H. GRAHAM  
Columbia University

ERNEST R. HILGARD  
Stanford University

HEINRICH KLÜVER  
University of Chicago

LEONARD MEAD  
Tufts University

NORMAN L. MUNN  
University of Adelaide,  
Australia

JOHN PAUL NAFE  
Florida State University

LEE EDWARD TRAVIS  
University of Southern  
California

ROBERT C. TRYON  
University of California

MICHAEL J. ZIGLER  
Wellesley College

JOSEPH ZUBIN  
Psychiatric Institute,  
Columbia University

---

## VOLUME 58

1964

---

Copyright, 1964, by The Journal Press  
Provincetown, Massachusetts, U. S. A.

All rights reserved

Published Quarterly at 35B New St., Worcester, Mass.

# AUTHOR INDEX

Abramson, H. A. ....	57, 237	Knehr, Charles A. ....	353
Aiken, Lewis R., Jr. ....	369	Krinsky, L. W. ....	237
Albee, George W. ....	141	Kubis, J. R. ....	459
Alexander, Theron ....	23	Kugel, Robert B. ....	23
Alluisi, Earl A. ....	223	Lane, Ellen A. ....	141
Badri, Malik B. ....	421	Lloyd, Kenneth E. ....	249
Baer, Daniel J. ....	357, 413, 467	Lorenz, Robert J. ....	353
Barber, Theodore Zenophon ....	215	Mackler, Bernard ....	205
Bartley, S. Howard ....	329, 333, 343, 379	Marston, Albert R. ....	169
Beecher, Henry K. ....	397	Mayzner, M. S. ....	3, 9, 17, 317
Bleck, Fred C. ....	343	McGinnies, Elliott ....	177
Bloch, M. ....	107	Meighan, Thomas W. ....	223
Braun, John R. ....	313	Morgan, C. J. ....	439
Bronzaft, A. ....	107	Moynihan, James F., S.J. ....	413
Brown, Donald W. ....	89	Nelson, Thomas M. ....	333, 343, 379
Calverley, David Smith ....	215	Palmer, Robert D. ....	157
Cameron, Jean S. ....	265	Phelan, Joseph G. ....	43
Cautela, Joseph ....	85	Pollack, Robert H. ....	145
Ceraso, John ....	65	Ranney, Jane E. ....	329
Coffee, Sandra ....	375	Reuter, Jeanette M. ....	141
Cohen, Arthur M. ....	115	Rigg, Melvin G. ....	427
Corning, W. C. ....	131	Rolo, A. ....	57, 237
Crannell, C. W. ....	295	Shontz, Franklin C. ....	205
Cushna, Bruce ....	23	Smith, Carroll E. ....	191
Cutter, Henry S. G. ....	361	Smith, Gene Marshall ....	397
Dennis, Wayne ....	421	Snadowsky, Alvin M. ....	33
Dube, C. Stuart, II ....	243	Snider, Bill ....	23
Eisenman, Russell ....	375	Stewart, Daniel K. ....	79
Gabriel, R. F. ....	3, 9, 17	Sutton, Rachel S. ....	407
Gassner, Suzanne M. ....	33	Tresselt, M. E. ....	317
Gettner, H. H. ....	57	Tuddenham, Read D. ....	307
Gold, Jerome ....	33	Vitro, Francis ....	85
Goldfarb, L. ....	237	Warm, Joel S. ....	243
Greenberg, Lewis F. ....	243	Welch, L. ....	107
Harper, Earl S. ....	333	Wendt, G. R. ....	265
Harper, Frank B. W. ....	187, 307	Williams, Robert ....	145
Hawkes, Glenn R. ....	223	Yamamoto, Kaoru ....	277, 285
Hayes, R. ....	107	Zagona, Salvatore V. ....	255
Howard, S. M. ....	459	Zurcher, Louis A., Jr. ....	255
Johnson, Ronald C. ....	301		



# TABLE OF CONTENTS

Short-term retention of paired-associate material . . . . .	3
BY M. S. MAYZNER AND R. F. GABRIEL	
Immediate recall for digram letter matrices . . . . .	9
BY M. S. MAYZNER AND R. F. GABRIEL	
The effect of spatial organization of the stimulus on short-term retention . . . . .	17
BY M. S. MAYZNER AND R. F. GABRIEL	
Studies of complex behavior: 1. The processes of perception, association, and prediction of response . . . . .	23
BY THERON ALEXANDER, ROBERT B. KUGEL, BRUCE CUSHNA, AND BILL SNIDER	
Changes in the phenomenal field as a result of human relations training . . . . .	33
BY SUZANNE M. GASSNER, JEROME GOLD, AND ALVIN M. SNADOWSKY	
Analysis of the "good errors" made by clinical psychologists in diagnostic judgment . . . . .	43
BY JOSEPH G. PHELAN	
Lysergic acid diethylamide (LSD-25): XXXV. Comparison of effect on Siamese fighting fish and goldfish . . . . .	57
BY H. H. GETTNER, A. ROLO, AND H. A. ABRAMSON	
Specific interference in retroactive inhibition . . . . .	65
BY JOHN CERAISO	
Communication and logic: VI. Evidence for showing the influence of negative thoughts on communication . . . . .	79
BY DANIEL K. STEWART	
The effect of instruction on the appearance of the autokinetic effect . . . . .	85
BY JOSEPH CAUTELA AND FRANCIS VITRO	
Two-color mixtures: I. Broad-band filters . . . . .	89
BY DONALD W. BROWN	
Intra-series primacy and the orienting reflex of the GSR . . . . .	107
BY R. HAYES, A. BRONZAFI, M. BLOCH, AND L. WELCH	
Predicting organization in changed communication networks: III. . . . .	115
BY ARTHUR M. COHEN	
Evidence of right-left discrimination in planarians . . . . .	131
BY W. C. CORNING	
Childhood intelligence of future schizophrenics and neighborhood peers . . . . .	141
BY GEORGE W. ALBEE, ELLEN A. LANE, AND JEANETTE M. REUTER	
Some nonpsychological variables in therapy defection in a child-guidance clinic . . . . .	145
BY ROBERT WILLIAMS AND ROBERT H. POLLACK	
Cerebral dominance and auditory asymmetry . . . . .	157
BY ROBERT D. PALMER	
Personality variables related to self-reinforcement . . . . .	169
BY ALBERT R. MARSTON	

Attitudes toward civil liberties among Japanese and American university students . . . . .	177
BY ELLIOTT MCGINNIES	
The California Psychological Inventory as a predictor of yielding behavior in women . . . . .	187
BY FRANK B. W. HARPER	
The effect of anxiety on the performance and attitudes of authoritarians in a small group situation . . . . .	191
BY CARROLL E. SMITH	
Life styles and creativity: A review . . . . .	205
BY BERNARD MACKLER AND FRANKLIN C. SHONTZ	
Hypnotizability, suggestibility, and personality: I. Two studies with the Edwards Personal Preference Schedule, the Jourard Self-Disclosure Scale, and the Marlowe-Crowne Social Desirability Scale . . . . .	215
BY THEODORE XENOPHON BARBER AND DAVID SMITH CALVERLEY	
Vigilance in complex task situations . . . . .	223
BY GLENN R. HAWKES, THOMAS W. MEIGHAN, AND EARL A. ALLUISI	
Multitherapist interviews utilizing LSD . . . . .	237
BY A. ROLO, L. W. KRINSKY, H. A. ABRAMSON, AND L. GOLDFARB	
Stimulus and motivational determinants in temporal perception . . . . .	243
BY JOEL S. WARM, LEWIS F. GREENBERG, AND C. STUART DUBE, II	
Short-term retention as a function of item frequency . . . . .	249
BY KENNETH E. LLOYD	
Participation, interaction and role behavior in groups selected from the extremes of the open-closed cognitive continuum . . . . .	255
BY SALVATORE V. ZAGONA AND LOUIS A. ZURCHER, JR.	
Chemical studies of behavior: VIII. Some anomalous onset effects of benzquinamide, a "tranquilizer" . . . . .	265
BY JEAN S. CAMERON AND G. R. WENDT	
A further analysis of the role of creative thinking in high-school achievement . . . . .	277
BY KAORU YAMAMOTO	
Evaluation of some creativity measures in a high school with peer nominations as criteria . . . . .	285
BY KAORU YAMAMOTO	
Code learning and color . . . . .	295
BY C. W. CRANNELL	
Mean associative latencies of 200 CVC trigrams . . . . .	301
BY RONALD C. JOHNSON	
The sociometric composition of the group as a determinant of yielding to a distorted norm . . . . .	307
BY FRANK B. W. HARPER AND READ D. TUDDENHAM	
Consistency of cross-test individual differences in personality-inventory faking . . . . .	313
BY JOHN R. BRAUN	



Shifts in connotative meaning of words as a function of context and reinforcement . . . . .	317
BY M. E. TRESSELT AND M. S. MAYZNER	
Critical flicker frequency for high pulse-to-cycle fractions and high intensity . . . . .	329
BY JANE E. RANNEY AND S. HOWARD BARTLEY	
CFF for short trains of photic stimulation having various temporal distributions and separations . . . . .	333
BY THOMAS M. NELSON, S. HOWARD BARTLEY, AND EARL S. HARPER	
The effects of cycling upon CFF when stimulus purity is varied . . . . .	343
BY THOMAS M. NELSON, S. HOWARD BARTLEY, AND FRED C. BLECK	
Subjective color: A new method of producing the phenomenon . . . . .	353
BY CHARLES A. KNEHR AND ROBERT J. LORENZ	
Trait judgments of military pilots from photographs . . . . .	357
BY D. J. BAER	
Conflict models, games, and drinking patterns . . . . .	361
BY HENRY S. G. CUTTER	
Item context and position effects on multiple-choice tests . . . . .	369
BY LEWIS R. AIKEN, JR.	
Aesthetic preferences of art students and mathematics students . . . . .	375
BY RUSSELL EISENMAN AND SANDRA COFFEE	
Description of a central mechanism related to diverse brightness phenomena . . . . .	379
BY THOMAS M. NELSON AND S. HOWARD BARTLEY	
Drugs and judgment: Effects of amphetamine and secobarbital on self-evaluation . . . . .	397
BY GENE MARSHALL SMITH AND HENRY K. BEECHER	
Behavior in the attainment of economic concepts: II. . . . .	407
BY RACHEL S. SUTTON	
Stepwise discriminant-function analysis of seminary-candidate MMPI scores . . . . .	413
BY DANIEL J. BAER AND JAMES F. MOYNIHAN, S.J.	
Human-figure drawings in relation to modernization in Sudan . . . . .	421
BY MALIK B. BADRI AND WAYNE DENNIS	
The mood effects of music: A comparison of data from four investigators . . . . .	427
BY MELVIN G. RIGG	
An individual action-research method for intensive exploration of adaptive difficulties in the prepuberty stage . . . . .	439
BY C. J. MORGAN	
Ego identity and some aspects of personal adjustment . . . . .	459
BY S. M. HOWARD AND J. F. KUBIS	
Set and shock-stress effects upon illusion perception . . . . .	467
BY DANIEL J. BAER	

\$20.00 per annum  
\$15.00 per volume  
\$7.50 per half volume

QUARTERLY  
Two volumes per year  
Immediate publication  
(Founded in 1935 by Carl Murchison)

July, 1964  
Volume 58, First Half

# The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

JULY, 1964

(Manuscripts are printed in the order of final acceptance)

Short-term retention of paired-associate material . . . . .	3
BY M. S. MAYZNER AND R. F. GABRIEL	
Immediate recall for digram letter matrices . . . . .	9
BY M. S. MAYZNER AND R. F. GABRIEL	
The effect of spatial organization of the stimulus on short-term retention . . . . .	17
BY M. S. MAYZNER AND R. F. GABRIEL	
Studies of complex behavior: 1. The processes of perception, association, and prediction of response . . . . .	23
BY THERON ALEXANDER, ROBERT B. KUGEL, BRUCE CUSHNA, AND BILL SNIDER	
Changes in the phenomenal field as a result of human relations training . . . . .	33
BY SUZANNE M. GASSNER, JEROME GOLD, AND ALVIN M. SNADOWSKY	

(OVER)

Copyright, 1964, by The Journal Press  
Provincetown, Massachusetts, U. S. A.

All rights reserved

Published at 35B New St., Worcester, Mass.  
Second-class postage paid at Worcester, Mass.



Analysis of the "good errors" made by clinical psychologists in diagnostic judgment By JOSEPH G. PHELAN	43
Lysergic acid diethylamide (LSD-25): XXXV. Comparison of effect on Siamese fighting fish and goldfish By H. H. GETTNER, A. ROLO, AND H. A. ABRAMSON	57
Specific interference in retroactive inhibition By JOHN CERASO	65
Communication and logic: VI. Evidence for showing the influence of negative thoughts on communication By DANIEL K. STEWART	79
The effect of instruction on the appearance of the autokinetic effect By JOSEPH CAUTELA AND FRANCIS VITRO	85
Two-color mixtures: I. Broad-band filters By DONALD W. BROWN	89
Intra-series primacy and the orienting reflex of the GSR By R. HAYES, A. BRONZAFI, M. BLOCH, AND L. WELCH	107
Predicting organization in changed communication networks: III. By ARTHUR M. COHEN	115
Evidence of right-left discrimination in planarians By W. C. CORNING	131
Childhood intelligence of future schizophrenics and neighborhood peers By GEORGE W. ALBEE, ELLEN A. LANE, AND JEANETTE M. REUTER	141
Some nonpsychological variables in therapy defection in a child-guidance clinic By ROBERT WILLIAMS AND ROBERT H. POLLACK	145
Cerebral dominance and auditory asymmetry By ROBERT D. PALMER	157
Personality variables related to self-reinforcement By ALBERT R. MARSTON	169
Attitudes toward civil liberties among Japanese and American university students By ELLIOTT MCGINNIES	177
The California Psychological Inventory as a predictor of yielding behavior in women By FRANK B. W. HARPER	187
The effect of anxiety on the performance and attitudes of authoritarians in a small group situation By CARROLL E. SMITH	191
Life styles and creativity: A review By BERNARD MACKLER AND FRANKLIN C. SHONTZ	205
Hypnotizability, suggestibility, and personality: I. Two studies with the Edwards Personal Preference Schedule, the Jourard Self-Disclosure Scale, and the Marlowe-Crowne Social Desirability Scale By THEODORE XENOPHON BARBER AND DAVID SMITH CALVERLEY	215
Vigilance in complex task situations By GLENN R. HAWKES, THOMAS W. MEIGHAN, AND EARL A. ALLUISI	223
Multitherapist interviews utilizing LSD By A. ROLO, L. W. KRINSKY, H. A. ABRAMSON, AND L. GOLDFARB	237

## SHORT-TERM RETENTION OF PAIRED- ASSOCIATE MATERIAL\*<sup>1</sup>

*New York University and Douglas Aircraft Corporation*

---

M. S. MAYZNER AND R. F. GABRIEL

---

### A. INTRODUCTION

In many of the current large computerized man-machine systems now being employed by the military services, displays are employed which present status information of various kinds. For example, such status displays might present a list of air bases and, associated with each air base name, a number indicating how many interceptors or missiles are available for immediate deployment, or a list of cities might be presented with associated numbers indicating target priority values. Such information typically is presented for short periods of time (e.g., 5-10 seconds) and then erased to allow other information items to appear on the display. At some subsequent time, it is often necessary to utilize the information previously displayed in order to make some decision or take some action, and it then becomes necessary either to recall the information or to have the display containing the information regenerated. The regeneration may take time or the display may be needed for the presentation of some other class of information items at this particular time. When these conditions occur, retention by the operator of the necessary information may be a critical factor in the success of a particular task or of an entire mission. Therefore, storage capacity and the parameters affecting such capacity need to be understood thoroughly to insure optimum design of such information displays.

The present study examines certain features of this type of information storage and retrieval problem. Specifically, Ss were presented with a list of cities and associated two-digit numbers, which they were informed represented the air temperatures of those cities on a particular day. Subsequently they were asked to recall the temperatures shown them. The two variables chosen for investigation were (a) the number of city-temperature pairs presented, which varied from three to six, and (b) the familiarity of the

---

\* Received in the Editorial Office on January 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by Contract Nonr 285(56) between the Engineering Psychology Branch of the Office of Naval Research and New York University.



city names, which was varied by choosing lists of cities with very large or very small populations.

## B. METHOD

### 1. *Subjects*

The *Ss* were 80 students from introductory psychology classes at the University of Southern California.

### 2. *Stimulus Material and Procedure*

The stimulus material consisted of 20 cities, each paired with a randomly selected two-digit temperature, with the constraint that no temperature repeated the same two digits or ended in a zero or five. Ten of the 20 cities had populations of 100,000 and over (Akron, Boston, Cleveland, Detroit, Erie, Houston, Memphis, New York, Pittsburgh, and Richmond), while the other 10 cities had populations of 10,000 and under (Belmar, Elwood, Freehold, Greeley, Harvel, Longmont, Malvern, Olney, Rahway and Tifton). The apparatus employed for presenting the city-temperature pairs was a Hunter Card Master. Each pair was exposed to *S* for a period of two seconds. Two seconds also elapsed between pairs, during which time the shutter is closed on the Card Master while a new stimulus card is being brought to the viewing window.

A  $2 \times 4$  analysis of variance design was employed, with 10 *Ss* assigned randomly to each of the eight conditions. The two variables examined were list length and city-name familiarity. Four levels of list length were chosen, employing three, four, five, or six city-temperature pairs, and two levels of city-name familiarity—high and low—which variable was manipulated by using cities with large populations (over 100,000) for high familiarity and cities with small populations (under 10,000) for low familiarity. Here the assumption is made that cities with large populations will be familiar to the average *S*, while those with small populations will not be familiar. Subsequent inquiry showed that all 80 *Ss* were quite familiar with the 10 large-population cities, while none of the 80 *Ss* had ever heard of the small-population cities.

As indicated, 10 large- and 10 small-population cities were selected and, for each *S*, depending on what condition he had been assigned, the set of 10 stimulus cards (large-population cities or small-population cities) was thoroughly mixed and three, four, five, or six cards were selected at random for display in the Card Master. Thus, each *S* received to some degree a different set of cities than any other *S*. This feature was employed



to randomize over the several conditions any biases due to particular city names. For the recall task, *S* was given an alphabetical list of either the 10 large- or the 10 small-population cities, depending on what condition (high or low familiarity) he had been assigned, and was asked to write in the associated temperature for those cities he had just seen in the viewing window of the Card Master. Thus, *S* was not required to recall the city names, only the associated temperatures, but the cities that were shown to him in the Card Master were always part of a larger set of 10 cities that appeared on his recall response form. This particular type of presentation and recall task, while somewhat different from the more classical paired-associate learning task, is very similar to many of the display and recall sequences found in current man-machine systems. All *Ss* were tested individually, and the number of correctly recalled temperatures was used as the dependent response measure.

### C. RESULTS AND DISCUSSION

Table 1 presents the mean number of temperatures recalled correctly for the four list lengths (three, four, five, and six) and the two levels of city populations (large and small). The application of a  $2 \times 4$  analysis

TABLE 1  
MEAN NUMBER OF TEMPERATURES RECALLED FOR FOUR LIST LENGTHS  
(THREE, FOUR, FIVE, AND SIX) AND TWO LEVELS  
OF CITY POPULATIONS (LARGE AND SMALL)

		List length			
		3	4	5	6
Large populations	<i>M</i>	2.10	2.10	1.90	2.70
	<i>SD</i>	.83	1.30	.94	1.27
Small populations	<i>M</i>	1.40	2.10	2.20	1.90
	<i>SD</i>	.92	1.14	.87	1.58

of variance gave *F* values of less than 1.00 for both main effects and the interaction term. From these results it would appear that the effects of the variables being studied are negligible. In many ways this result is quite surprising. Much evidence (3) exists showing the effects of stimulus familiarity or meaningfulness on paired-associate learning. In our task no effects appear, however, since our variable of familiarity is manipulated in a rather unique way (i.e., cities with large or small populations); this may account, in part, for the absence of any effects. Even more surprising, however, is the failure of list length to produce any effects. It had been expected that, with the list length of three pairs, almost perfect recall



would result. As Table 1 shows, this was not the case. For the large-population cities, the mean number of temperatures recalled was 2.1, and of the 10 *Ss* only four *Ss* had a perfect score of 3, while three *Ss* had a score of 2 and three *Ss* had a score of 1. For the small-population cities, the mean number of temperatures recalled was only 1.4, and of the 10 *Ss*, only one *S* had a perfect score of 3, while four *Ss* had a score of 2, three *Ss* had a score of 1, and two *Ss* had a score of zero. This abysmal performance is not only difficult to understand but, if it is generally true, this type of task has some very unfortunate implications for display design vis-à-vis system design and man's information storage capacity for this particular type of material. Further, as list length increases, no systematic increase or decrease occurs in recall performance. Over the range examined, from three to six pairs, it appears as though *S* had a constant and very small storage capacity of approximately two temperatures or four digits. Being presented with more or less information has little effect on his capacity.

The very poor storage capacity found in this study finds some support in the recent work of Murdock (1, 2). In this research he displayed pairs of high-frequency English words for a single presentation and tested for recall of certain critical pairs by presenting the stimulus term to *S* and asking him to supply the response term. Murdock tested a wide variety of conditions in a series of experiments and states, "The present results clearly indicate that, under the conditions of these experiments, there is very rapid forgetting of single paired associates" (2, p. 441). Murdock's findings are in good agreement with our own, but the puzzle of why performance is so poor for this type of task remains.

Since storage limitations appear so marked with paired-associate-type tasks, it is strongly recommended that display designers be particularly sensitive to these limitations in their system-design efforts. Additional research is certainly indicated to explore more fully the parameters affecting these limitations. For example, display timing factors may be very critical in setting the boundaries on storage capacity, and parametric study of varying exposure times for the stimulus-response pair, coupled with varying exposure times for between-pair presentations, is needed to determine if any display timing optimums exist which maximize storage capacity.

#### D. SUMMARY

The present study was designed to examine the effects of list length and stimulus familiarity on the short-term retention of paired-associate

material. A  $2 \times 4$  analysis-of-variance design was employed with four levels of list length (three, four, five, and six stimulus-response pairs) and two levels of stimulus familiarity (high and low). The results indicated that both variables had negligible effects, and it was concluded that storage capacity for paired-associated material is very poor.

## REFERENCES

1. MURDOCK, B. B., JR. Short-term retention of single paired associates. *Psychol. Rep.*, 1961, **8**, 280.
2. ———. Short-term retention of single paired associates. *J. Exper. Psychol.*, 1963, **65**, 433-443.
3. UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and Verbal Learning*. Philadelphia: Lippincott, 1960.

*School of Engineering and Science*  
*New York University*  
*University Heights*  
*Bronx 53, New York*



## IMMEDIATE RECALL FOR DIGRAM LETTER MATRICES\*<sup>1</sup>

*New York University and Douglas Aircraft Corporation*

---

M. S. MAYZNER AND R. F. GABRIEL

---

### A. INTRODUCTION

In a previous study (1), it was indicated that the development of large-scale computerized man-machine systems has produced an increasing concern with man's capacity to process quickly and accurately large quantities of information and that a critical component of such information processing is man's capacity for information storage and retrieval. The previous study examined the effects of one parameter believed to be important to such storage and retrieval processes in man (namely, the effects of information organization or "chunking") and it was found that the organization of the stimulus material has a significant effect on the amount retained and subsequently recalled.

The present study is concerned with examining another parameter believed important to information storage and retrieval—namely, the coding of the information being presented to *S*. For example, in many man-machine systems the information displayed to the operator is presented in an "alpha" or "alpha-numeric" code; i.e., as an array of letters or letters and numbers. The use of a pure "alpha" code immediately raises the question of what letters or letter combinations to select, if retention of the information is critical. Specifically, if the information of whatever sort is coded into two-letter combinations or digrams and the display consists of a matrix of such digrams (such displays being quite common in many current systems), are certain digrams "better" to use than others; i.e., are certain digrams retained better than others? The present study seeks a preliminary answer to this question.

### B. METHODS

#### 1. *Subjects*

The *Ss* were 80 students from introductory psychology classes at the University of Southern California.

---

\* Received in the Editorial Office on January 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by Contract Nonr 285(56) between the Engineering Psychology Branch of the Office of Naval Research and New York University.

## 2. *Stimulus Material and Procedure*

The stimulus material consisted of two sets of 16 digrams. These digrams were taken from tables given by Underwood and Schulz (3) and were selected to provide one set of digrams with high frequency values and one set of digrams with low frequency values, as based on a 15,000-word sample. The high frequency digrams were ED, NT, NG, TE, OU, EA, EN, LE, TH, VE, HI, ES, AR, ST, ND, and AL, and ranged in frequency value from a low of 448 to a high of 1795, with a median value of 602. The low frequency digrams were EH, LT, TN, SD, AE, OE, HS, NR, NL, TN, EA, EU, DT, VA, EG, and EI, and ranged in frequency value from a low of 1 to a high of 465, with a median value of 15. Both sets of digrams consisted of the same identical pool of 32 letters. Each set was presented to *S* in a  $4 \times 4$  or 16-cell matrix, each cell containing a single digram, and the order of digrams from left to right for the four rows of each matrix was the same as that given above. These particular orders were chosen so that the frequency of the digrams formed by the second letter in each cell and the first letter in the next adjacent cell (at the end of a row the next adjacent letter is the first letter in the cell of the next row) would be approximately the same for the high and the low digram frequency sets. For example, for the high frequency digram matrix, the between-cell digram frequency values range from 1 to 242, with a median value of 38, while for the low frequency digram matrix the between-cell digram frequency values range from zero to 447, with a median value of 38. Thus, both digram matrices contain the same pool of 32 letters and vary only in the within-cell digram frequency values, not in the between-cells digram frequency values.

Eight groups of 10 *Ss* each were employed in a  $2 \times 4$  analysis-of-variance design, with two digram frequency values, high and low, and four matrix exposure times, 15, 30, 60, and 120 seconds. *Ss* were assigned randomly to the eight different experimental conditions. Each matrix was arranged in a  $4 \times 4$  cell array on white  $6 \times 6$  inch cards, using  $\frac{1}{4}$ -inch-high "Letraset" sheet No. 129 letters. *Ss* were instructed to study the array while it was being exposed, and immediately after the exposure duration period of either 15, 30, 60, or 120 seconds was completed, *S* was presented with a blank  $4 \times 4$  cell matrix and asked to fill in as many of the cells as he could recall, with the letter pairs just shown to him. The number of cells correctly filled (i.e., the number of digrams correctly recalled, and placed in the appropriate cell) was employed as the dependent response measure, to examine the effects of digram frequency and exposure time.



## C. RESULTS AND DISCUSSION

Table 1 gives the mean number of digrams recalled correctly for the four exposure times (15, 30, 60, and 120 seconds) and the two levels of digram frequency (high and low). The application of Barlett's test indicated heterogeneity of variance, therefore a square-root transformation of the data was performed, yielding a satisfactory value indicating homogeneity of variance (chi square = 12.16;  $df = 7$ ;  $p > .05$ ). The  $2 \times 4$  analysis of variance on the transformed scores indicated that both main effects were significant: For exposure time,  $F = 12.38$ ;  $df = 3, 71$ ;  $p < .001$ . For digram frequency,  $F = 5.83$ ;  $df = 1, 71$ ;  $p < .02$ , while the interaction term was insignificant ( $F = .74$ ;  $df = 3, 71$ ;  $p > .20$ ).

TABLE 1  
MEAN NUMBER OF DIGRAMS RECALLED FOR FOUR EXPOSURE TIMES  
(15, 30, 60, AND 120 SECONDS) AND TWO LEVELS  
OF DIGRAM FREQUENCY (HIGH AND LOW)

Frequencies	Exposure time (seconds)			
	15	30	60	120
Low digram				
<i>M</i>	1.60	3.70	5.00	7.50
<i>SD</i>	.92	1.27	2.49	3.50
High digram				
<i>M</i>	3.60	5.60	5.90	8.10
<i>SD</i>	1.28	2.15	3.81	3.91

The interpretation of the results shown in Table 1 appears quite straightforward; as exposure time increases from 15 seconds to 120 seconds, recall performance increases, as would be expected. Further, for each value of exposure time, more high frequency digrams are recalled than low frequency digrams. This latter finding is in agreement with the general findings of Underwood and Schulz (3) that high frequency digrams typically are learned faster and retained better than low frequency digrams, and extends the generality of Underwood and Schulz's work to a type of recall task very different from any employed in their research. It is interesting that—while the interaction term was not significant—as exposure time increases, the differences between high and low digram frequency conditions decrease from a value of 2.00 (3.60 — 1.60) to a value of .60 (8.10 — 7.50). This suggests that in this type of task, where all stimuli are presented simultaneously for a period of study and then recall is tested, the effects of letter or digram frequency are strongest for short exposure times. This type of finding may be related to word threshold studies (2), where usually low frequency words have higher thresholds than high frequency words,

and possibly indicates a certain functional similarity between perceptual threshold tasks and recall tasks for simultaneously presented stimuli (i.e., short exposure times are required to show the effect of frequency factors on perceptual thresholds and on recall scores).

Since differences in between-cell digram frequency values were equated for the high and the low digram frequency conditions, the significant differences obtained between the high and the low frequency conditions may be attributed to the within-cell digram frequency differences. However, this still leaves unanswered the question of whether between-cell differences might also produce differences in recall performance. Therefore, a second experiment was performed to answer this question.

Two  $4 \times 4$  digram matrices were constructed employing the same 16 digrams in both matrices, insuring that the within-cell digram frequencies would be identical for the two matrices, and were administered to two groups of 20 Ss each for a period of 120 seconds. The order of the 16 digrams in the matrices was made different, however, yielding a high between-cell digram frequency matrix and a low between-cell digram frequency matrix. The digrams in order from left to right for the four rows for the high between-cell matrix were as follows: EA, BL, DR, NC, HT, OS, IG, HO, MI, VE, LY, OP, AR, SH, AI, and CT. The between-cell frequency values for this matrix ranged from a low of 107 to a high of 558, with a median value of 245. The digrams in order from left to right for the four rows for the low between-cell matrix were as follows: EA, OP, BL, NC, DR, HO, AR, HT, VE, OS, LY, AI, IG, SH, CT, and MI. The between-cell frequency values for this matrix ranged from a low of zero to a high of 43, with a median value of 8.

TABLE 2  
MEAN NUMBER OF DIGRAMS RECALLED FOR HIGH  
AND LOW BETWEEN-CELL DIGRAM FREQUENCY CONDITIONS

Frequency	High	Low
<i>M</i>	7.95	6.85
<i>SD</i>	4.00	4.48

The mean number of digrams recalled for these high and low between-cell digram frequency conditions are presented in Table 2. The results show better recall for the high between-cell condition than for the low between-cell condition, but the difference of 1.10 is not significant ( $t = .80$ ;  $df = 38$ ;  $p > .40$ ). Thus, it would appear that the between-cell digram frequency difference, while in the same direction as the within-cell difference, is not as strong as the within-cell difference in producing effects on recall performance.



However, since the between-cell difference is in the same direction as the within-cell difference, it may well be a source of variance, nevertheless, in tasks such as these (particularly at shorter exposure times, where the within-cell difference has its strongest effect).

Of some interest in a recall task of the type employed in this study are the recall scores for each of the separate 16 cells of the matrix. Such an analysis was performed for each of the eight conditions of the first experiment and for each of the two conditions of the second experiment. Since the pattern was quite similar for all 10 conditions, the data were pooled and are presented in a combined form for all 10 conditions in Table 3.

TABLE 3  
TOTAL NUMBER OF DIGRAMS RECALLED FOR EACH CELL SUMMED  
ACROSS THE TEN EXPERIMENTAL CONDITIONS

Rows	Columns				Subtotals
	1	2	3	4	
1	84	49	50	40	223
2	53	43	26	33	155
3	55	26	35	40	156
4	45	45	29	53	172
Subtotals	237	163	140	166	706

Table 3 shows that the grand total for all 16 cells combined and summed over the 10 conditions is 706 digrams correctly recalled. If 706 is divided by 16, a value of approximately 44 is obtained, which represents the average number of digrams we might expect to be recalled in each cell if a uniform distribution of recall is obtained across the entire matrix. The results of Table 3 strongly indicate that a uniform distribution of recall is not found. Examination of Table 3 reveals rather that the cells in column 1 and row 1 are definitely favored as well as the last cell in row 4. Such a pattern is probably functionally equivalent to the serial position effect of serial learning tasks, but it is interesting to note that in the matrix type of task both the first row and the first column receive favored attention. This is somewhat unexpected, because reading habits would predict superior recall for row 1 but not for column 1. Apparently, however, some of the *Ss* in this study attend as much to column 1 as they do to row 1, or another possibility would suggest that about half of the *Ss* attend principally to row 1 and the other half to column 1. In any case, the finding is intriguing and deserves further investigation with a wider range of stimulus materials and matrix sizes.

Related to the distribution of recall scores for the separate cells of the matrix is the general issue of acquisition and recall strategies. Here reference

is made to the information processing that is carried out by *S* during the exposure period, when he is examining the matrix, and subsequently during the recall period, when he is attempting to fill in the cells of an empty matrix.

During the acquisition phase, as revealed by questioning of the *Ss* at the completion of the experiment, *Ss* employ a large variety of devices to encode the information presented to them. For example, some *Ss* report attempting to construct partial sentences using the digrams as cues, other *Ss* look for initials of friends and objects, others try to construct a visual image, and approximately half of the *Ss* report attempting to associate the digrams to words. With such a rich variety of encoding strategies, an immediate question arises as to whether certain encoding strategies are superior to others, and this most probably is a function of the stimulus material itself.

Recall strategies also differ, with some *Ss* filling in the blank matrix in a systematically horizontal fashion, proceeding from row to row in an orderly manner; others fill in the matrix in a systematically vertical fashion, proceeding from column to column in an orderly manner; and still others fill in the matrix in a mixed fashion, jumping seemingly at random from cell to cell in a zig-zag pattern. Certainly the present study does not clarify any of these complex issues. It does indicate, however, that a systematic treatment of acquisition and recall strategies is particularly necessary for the type of task employed in this study.

In brief, the results of the present study strongly suggest that in any man-machine display design in which information storage and retrieval may be a factor, and in which the display employs an "alpha"-type code, letter or digram frequencies can provide an important source of variance in recall performance. It is recommended that, whenever possible in such display designs, digrams with high frequencies of occurrence in the language should be used to capitalize on their superior storage and retrieval properties.

#### D. SUMMARY

The present study was designed to examine the effects of digram frequencies on short-term retention. *Ss* were presented with a  $4 \times 4$  display matrix, with each of the 16 cells containing a two-letter pair or digram, and after a period of study were asked for their immediate recall of the matrix. A  $2 \times 4$  analysis of variance design was employed with four levels of exposure time (15, 30, 60, and 120 seconds) and two levels of digram frequency (high and low). The results showed a systematic increase in recall scores with increasing exposure time for both levels of digram frequency and superior recall of high



frequency digrams over low frequency digrams for each level of exposure time. It was recommended that high frequency digrams be employed whenever possible in the design of man-machine displays employing an "alpha"-type code.

#### REFERENCES

1. MAYZNER, M. S., & GABRIEL, R. F. Information "chunking" and short-term retention. *J. of Psychol.*, 1963, **56**, 161-164.
2. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.
3. UNDERWOOD, B. J., & SCHULZ, R. W. Meaningfulness and Verbal Learning. Philadelphia: Lippincott, 1960.

*School of Engineering and Science*  
*New York University*  
*University Heights*  
*Bronx 53, New York*

## THE EFFECT OF SPATIAL ORGANIZATION OF THE STIMULUS ON SHORT-TERM RETENTION\*<sup>1</sup>

*New York University and Douglas Aircraft Corporation*

---

M. S. MAYZNER AND R. F. GABRIEL

---

### A. INTRODUCTION

In a previous study (2) the effect of information organization or "chunking" on short-term retention was examined. Specifically, five different groups of 10 Ss each received the same set of 12 digits for a period of six seconds, but each group received the 12 digits in a different organizational array; i.e., 12 lines (one digit per line), six lines (two digits per line), four lines (three digits per line), three lines (four digits per line), and two lines (six digits per line). The results showed a systematic increase in retention from 12 lines to two lines.

The present study is concerned with examining this spatial organization parameter further with a different type of stimulus material—namely, letter pairs or digrams. Since an earlier study (3) employed a  $4 \times 4$  cell matrix of either a set of 16 high or low frequency digrams, with one digram per cell, the present experiment extends this previous work by presenting the set of 16 high frequency digrams in a  $2 \times 8$  array, an  $8 \times 2$  array, and a circular array. The question posed is "will retention vary as a function of the spatial organization of the stimulus?"

### B. METHOD

#### 1. Subjects

The Ss were 90 students from introductory psychology classes at the University of Southern California. None of the Ss had participated in the earlier digram study (3).

#### 2. Stimulus Material and Procedure

The stimulus material consisted of a set of 16 high frequency digrams taken from tables given by Underwood and Schulz (4) and were the same

---

\* Received in the Editorial Office on January 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by Contract Nonr 285(56) between the Engineering Psychology Branch of the Office of Naval Research and New York University.



16 high frequency digrams employed in the earlier-mentioned study (3). The 16 digrams were as follows: ED, NT, NG, TE, OU, EA, EN, LE, TH, VE, HI, ES, AR, ST, ND, and AL. They ranged in frequency value from a low of 448 to a high of 1795, with a median frequency value of 602.

Three stimulus arrays were prepared. The first array consisted of the 16 high frequency digrams placed on a  $10 \times 3\frac{1}{2}$  inch plain white card in a  $2 \times 8$  configuration (i.e., two horizontal rows of eight digrams each). The second array of the same 16 digrams was placed on a  $3\frac{1}{2} \times 10$  inch plain white card in an  $8 \times 2$  configuration (i.e., eight horizontal rows of two digrams each). The third array of the same 16 digrams was placed on a  $7 \times 7$  inch plain white card in a circular pattern. The letters were  $\frac{3}{4}$ -inch-high "Letraset" sheet No. 129 letters. The  $2 \times 8$  and  $8 \times 2$  arrays presented the digrams in the same order as given above, reading from left to right across rows. And in the circular pattern, the twelve-o'clock position was ED; the one-o'clock position, NT; the two-o'clock position NG, etc. Ss were presented the cards for a period of 120 seconds and then were asked for their immediate recall, by filling in the blank cells of a  $2 \times 8$ , and  $8 \times 2$ , or a circular-array response form. Forty-five Ss were assigned randomly to each of the three stimulus array conditions (15 Ss per condition), but since the results of this experiment were not significant, the experiment was replicated with a second group of 45 Ss. The results from the second group were not significantly different from those for the first group and, therefore, and combined results of both groups ( $N = 90$  Ss) are presented in the Results Section. The number of cells correctly filled on the response forms—i.e., the number of digrams correctly recalled—was employed as the dependent measure, to examine the effects of spatial organization of the stimulus on retention.

### C. RESULTS AND DISCUSSION

Table 1 presents the mean number of digrams recalled correctly for the  $2 \times 8$ , the  $8 \times 2$ , and the circular array. An application of Barlett's test indicated homogeneity of variance, and therefore a simple analysis of variance

TABLE 1  
MEAN NUMBER OF DIGRAMS RECALLED FOR THE  $2 \times 8$ ,  
THE  $8 \times 2$ , AND THE CIRCULAR ARRAY

	$2 \times 8$	Stimulus arrays	
		$8 \times 2$	Circular
<i>M</i>	8.97	7.17	8.66
<i>SD</i>	3.26	2.94	3.63

was carried out yielding an  $F$  value of 2.71 ( $df = 2, 87$ ) and a  $p$  value between .05 and .10. Since this analysis was performed on the combined data from two separate replications, neither of which alone was significant, and since the combined analysis also fails to reach the .05 level of significance, the evidence is meager for concluding that an effect of spatial organization on retention exists. However, since the  $p$  value does fall between the .05 and .10 levels, the results are suggestive that the spatial organization of the stimulus material does have an effect on retention, but it is a weak effect. The relative ordering of the three stimulus arrays with respect to recall performance is generally in line with reading habits, at least with regard to the superiority of the  $2 \times 8$  array over the  $8 \times 2$  array, but a more precise predictive ordering of recall performance for a variety of stimulus organizations must await future theoretical development in this area.

An analysis was made also of the total number of digrams recalled correctly for each of the 16 cells in the three different stimulus arrays, and the results are presented in Table 2. In each of the three arrays, a

TABLE 2  
TOTAL NUMBER OF DIGRAMS RECALLED FOR EACH CELL  
IN  $2 \times 8$ , THE  $8 \times 2$ , AND THE CIRCULAR ARRAY

<i>2 × 8 Array</i>								
Row 1	26	18	18	18	17	8	17	20
Row 2	26	15	20	7	8	12	12	27
<i>8 × 2 Array</i>								
Row 1	30	25						
2	15	14						
3	20	16						
4	8	8						
5	10	8						
6	10	10						
7	11	11						
8	10	9						
			<i>Circular Array</i>					
			25					
			15 23					
			9 20					
			16 10					
			20 24					
			12 7					
			16 8					
			18 13					
			24					

uniform distribution of recall scores over the 16 cells does not seem to hold, but rather certain cells seem to be favored. For example, in the  $2 \times 8$  array the cells in the first three columns and the last column are favored over columns four through seven. In the  $8 \times 2$  array the first row and, to a lesser extent, the second and third rows are favored over the last five rows. In the circular array the twelve-, three-, six-, and nine-o'clock positions seem favored somewhat over the intermediate positions.

In an earlier study (3) the 16 digrams of the present study were pre-



sented in a  $4 \times 4$  matrix array, and it is interesting to note that with the  $4 \times 4$  array the mean number of digrams recalled was 8.10, a value approximately midway between the 7.17 digrams recalled with the  $8 \times 2$  array and 8.97 digrams recalled with the  $2 \times 8$  array. Also, with the  $4 \times 4$  matrix array, certain cells (i.e., the first row, the first column, and the last cell in the last row) were favored over others with respect to recall performance.

Thus, the results from both studies taken together provide tentative support for our hypothesis that the spatial organization of the stimulus array does, indeed, affect recall performance not only with respect to the entire stimulus array, but differentially; that is, certain cells are favored over others in recall, depending on the particular array under consideration.

Additional support for our hypothesis concerning the effects of the spatial organization of the stimulus may be found in a study by Asch, Hay, and Diamond (1). In this study an identical list of 12 three-letter nonsense syllables was presented in a linear array of 12 positions, along a triangular contour, and alternately in two adjacent positions, using the method of anticipation. The results indicated that these figural or spatial conditions produced significant differences in the rate of learning. For example, the triangular organization was easiest, the two alternating positions the hardest. Using a different method then (i.e., the method of anticipation) rather than a simultaneous exposure of the entire stimulus array and different spatial configurations, as well as trigrams rather than digrams, Asch, Hay, and Diamond also found effects due to spatial organization.

The conclusion seems fairly clear—on the basis of the present study, the Asch, Hay, and Diamond study, and our earlier study (2) employing digits as the stimulus materials—that the spatial organization of the stimulus can and does exert effects on learning and recall performance. The complex issue of how spatial organization transforms learning and recall performance remains a problem for further theorizing and research. Certainly the acquisition and recall strategies outlined in our previous study on the recall of letter digram matrices (3) would play an important role in such an endeavor and possibly represent an initial point of departure for attacking this problem.

#### D. SUMMARY

The present study was designed to examine the effects of spatial organization of the stimulus on short-term retention. Three different groups of 30 Ss each received the same set of 16 high frequency digrams arranged in a  $2 \times 8$ , an  $8 \times 2$ , and a circular array for a period of 120 seconds, and

then were asked for their immediate recall. The results showed recall to be best for the  $2 \times 8$  array, next for the circular array, and poorest for the  $8 \times 2$  array; however, the differences fell just short of significance at the .05 level. It was concluded that these results, taken together with the results from three other studies, do indicate that the spatial organization of the stimulus can significantly affect short-term retention.

## REFERENCES

1. ASCH, S. E., HAY, J., & DIAMOND, R. M. Perceptual organization in serial rote-learning. *Amer. J. Psychol.*, 1960, **73**, 177-198.
2. MAYZNER, M. S., & GABRIEL, R. F. Information "chunking" in short-term retention. *J. of Psychol.*, 1963, **56**, 161-164.
3. ———. Immediate recall for digram letter matrices. *J. of Psychol.*, 1964, **58**, 9-15.
4. UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and Verbal Learning*. Philadelphia: Lippincott, 1960.

*School of Engineering and Science*  
*New York University*  
*University Heights*  
*Bronx 53, New York*



## STUDIES OF COMPLEX BEHAVIOR: I. THE PROCESSES OF PERCEPTION, ASSOCIATION, AND PREDICTION OF RESPONSE\*<sup>1</sup>

*Department of Pediatrics, University Hospitals, State University of Iowa*

---

THERON ALEXANDER, ROBERT B. KUGEL, BRUCE CUSHNA, AND  
BILL SNIDER

---

### A. INTRODUCTION

The study described here is the first of three reports covering research carried on over a number of years (1, 2) to find the characteristics which impair individuals in interaction with their environment. Too often attention is directed to the question of whether or not an individual is impaired and a description of the significance of his specific kinds of impairment is neglected. For example, sometimes individuals are described as having central-nervous-system disorder (CNS), or brain aberration, which may also cause the term "mental retardation" to be used, and yet it is difficult to describe adequately wherein the individual fails to behave effectively. Similarly, diagnosticians too often have used the category of "behavior disorder" or "emotional disorder" and have not described the ways in which the difficulty interferes with the individual's effectiveness in environmental interaction. Accordingly, the investigators endeavored in this study to go beyond the mere placing of a label on the individual's ineffective behavior to differentiate between normality and pathology. An effort was made to study behavior from a broad perspective, including the whole process of the individual's dealing with a stimulus and his responding to it. Thus this approach is considered to be molar rather than molecular.

Environmental interaction for this study is viewed as consisting of awareness or perception of the present stimulus situation, use of experience, and the prediction of future events and the individual's response to them. Based

---

\* Received in the Editorial Office on February 3, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> We are grateful for the support of the University Computer Center in this research. We also appreciate the support of Donal Dunphy, M.D., Chairman of the Department of Pediatrics. Marjorie Clatterbaugh of the Iowa City public schools was of assistance to us in making the normal subjects available. We also acknowledge the contribution of our research assistants Tanya Bohlke, Ann Don Carlos, and Doris Palmer. Robert B. Kugel, M.D. is now on the faculty of Brown University.

on the above viewpoint, a continuum of complex behavior was constructed with five categories: enumeration, enumeration and description, causation, a restricted description of future events, and prediction of independent purposeful action.

The main hypothesis is that children with both types of pathology, CNS disorder and behavior disorder, will have their impairment reflected in their performance on the behavioral continuum. The fundamental purpose of this study thus is to find at what point on the continuum the groups characteristically fall.

This study was specifically limited. The investigators did not undertake to develop a technique for the determination of CNS disorder, because it is believed that such determination is better done with medical techniques. Further, the techniques employed here were not seen as associated with the usual concepts of "intelligence" and it was not expected that there would necessarily be a high correlation with scores on intelligence tests, or *IQ*. In fact, the investigators did not wish to involve such concepts in their theoretical approach, but desired to take a different direction. The theoretical concept of intelligence has grown out of performance on varied but usually brief behavioral sequences. The purpose here is to study a sequence of complex behavior going beyond the briefer types of items usually included in intelligence tests.

The theoretical approach used in this study is in contrast to the theoretical approach of intelligence tests. Intelligence tests are constructed on the basis that items most frequently mastered are characteristic of a particular age level and thereby characterize what is to be considered normal or optimum behavior for that age level. In other words, a child's behavior which is in keeping with that of his peers is defined as intelligent behavior. Hence, in reality no theoretical criteria are applied against this behavior in terms of evaluation other than agreement with the majority. The approach used here is one in which effective behavior is theoretically defined without reference to whether or not the majority attain it. The sequence was set forth on the continuum and then this measure was applied against a large group of children who were viewed as having no outstanding difficulties as well as against those who had been separated from their peers by concern about their behavior.

## B. METHOD

### 1. *Subjects*

The subjects for the investigation include a "normal" group of 373 children in an elementary school, 30 children diagnosed as having behavior



disorder but no outstanding medical difficulties, and 32 children with central-nervous-system disorder (encephalopathy). All groups were similar in age and sex characteristics. The age range is from 5 to 12 years. Table 1 presents the diagnostic categories of the groups.

TABLE 1  
DIAGNOSTIC CATEGORIES OF THE GROUPS

Group	N
Behavior Disorder	
Overinhibited	17
Negativistic	10
Asocial	3
Total Behavior Disorder	30
CNS Disorder Group	
Encephalopathy due to:	
Prenatal anoxia	4
Mechanical birth injury	1
Anoxemia at birth	7
Postnatal anoxemia	3
Unknown cause	17
Total CNS Disorder	32

## 2. Experimental Task and Procedure

The categories on the continuum of complex behavior were given numerical values from one to five as shown in Figure 1.

The entire experimental task is designated as the Behavioral Complexity Test (BCT) and consists of three parts. Part I consists of tasks the completion of which depends upon first perception of form, then color, and finally the use of symbolization or meaning—this type of information about the subjects is obtained through the use of formboards. In Part II, a more

Enumeration	Enumeration & Description	Discontinuity	Supplantation Continuity	Continuity
E	ED	D	SC	C
1	2	3	4	5

FIGURE 1  
POINTS AND VALUES OF THE BEHAVIORAL CONTINUUM

difficult part than the first, the subject can select appropriate blocks for formboards in order to complete related pictures with missing parts. Part III, a more difficult series of tasks than Part II, requires that the subject rely more on his own resources and must tell a story about five pictures. The three parts of the test are shown in Figure 2.

The relationship of the parts of the BCT consisting of the scores for each part was investigated by studying the scores of the normal group of children. The scores for each of the three parts were converted to standard scores and the significance of the differences between the mean scores of the parts was determined. Analysis of variance indicated significant differences among the parts of the test. Further investigation of the differences among the groups is illustrated in Table 2. Also in Table 2 it will be seen

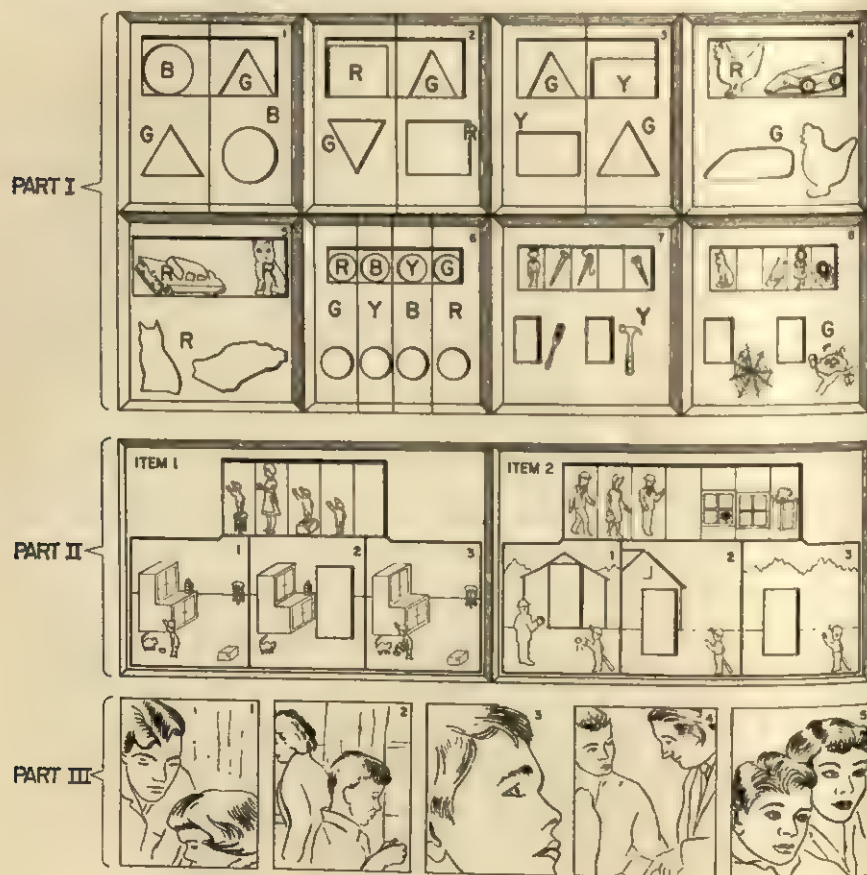


FIGURE 2

## SCHEMATIC DRAWING OF THE BEHAVIORAL COMPLEXITY TEST

Parts I and II consist of formboards. Subjects must place blocks in appropriate openings. The colors of the formboards in Part I are designated by letters: B = blue, G = Green, R = red, and Y = yellow. The formboards in Part II are green and the figures are in various colors. Part III consists of cards with black and white photographs about which subjects are asked to tell stories.



that the means decrease in size, thus indicating increasing difficulty of the parts (see also Figure 5).

Reliability was investigated through interscorer and test-retest correlations. For 45 subjects of the normal group the correlation between different scorers was .97; for the same 45 subjects the test-retest (approximately a three-month interval) correlation was .77.

TABLE 2  
STANDARD SCORE DIFFERENCE BETWEEN  
PARTS OF BCT CONTINUUM  
(NORMAL GROUP [ $N = 373$ ])

	Mean	SD	<i>t</i>
Part I	9.68	.88	
Part II	7.06	1.87	7.75*
Part I	9.68	.88	
Part III	5.90	1.30	46.61*
Part II	7.06	1.87	
Part III	5.90	1.30	3.12*

\* Significant at .01 level.

The BCT was administered individually to all subjects. Diagnoses of the clinical groups were reached through psychological study including testing, observation, and parent interviews, and through medical study including prenatal and developmental histories, physical examinations, routine laboratory studies, and electroencephalograms.

### C. RESULTS

The characteristics of the performance of the groups are shown in Figures 3 and 4. In general the subjects in the normal and behavior-disorder groups do not differ significantly on the continuum, but both of these groups differ from the CNS disorder group. This latter group has lower scores than the other two. Most of the scores of the children in the CNS disorder group as shown in Figure 4 fall at the second category (Enumeration and Description) while most of the scores of the normal and behavior-disorder children fall at the Discontinuity or "D" category. Although the continuum was constructed on a theoretical basis, in its use it was found that the scores of children without neurological impairment (both normal and behavior-disorder groups) fell in statistically normal distributions. The group differences on the continuum are shown in Table 3.

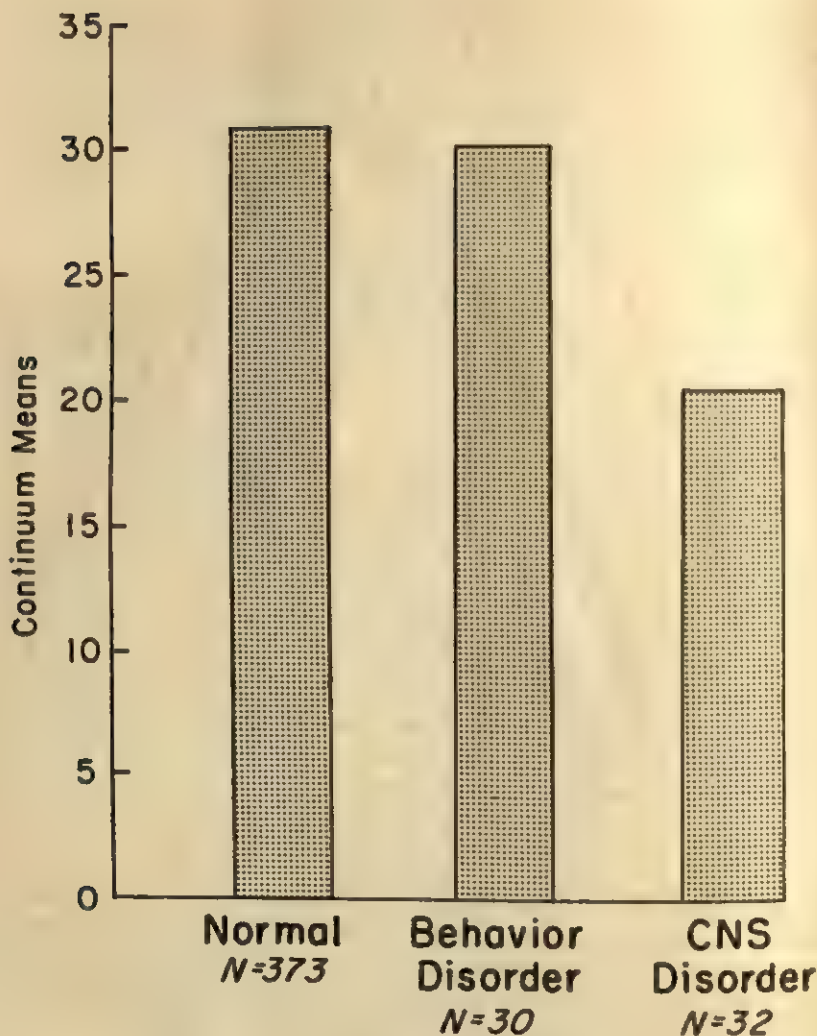


FIGURE 3  
CONTINUUM GROUP MEANS

1. *Age as a Factor*

The influence of age upon complex reasoning as shown by performance on the continuum was explored. The findings are shown in Tables 4 and 5. It will be noted that there is a positive correlation with age on the variables studies. Performance on the continuum, however, does not increase after about 8 or 9 years of age as illustrated in Figure 5.

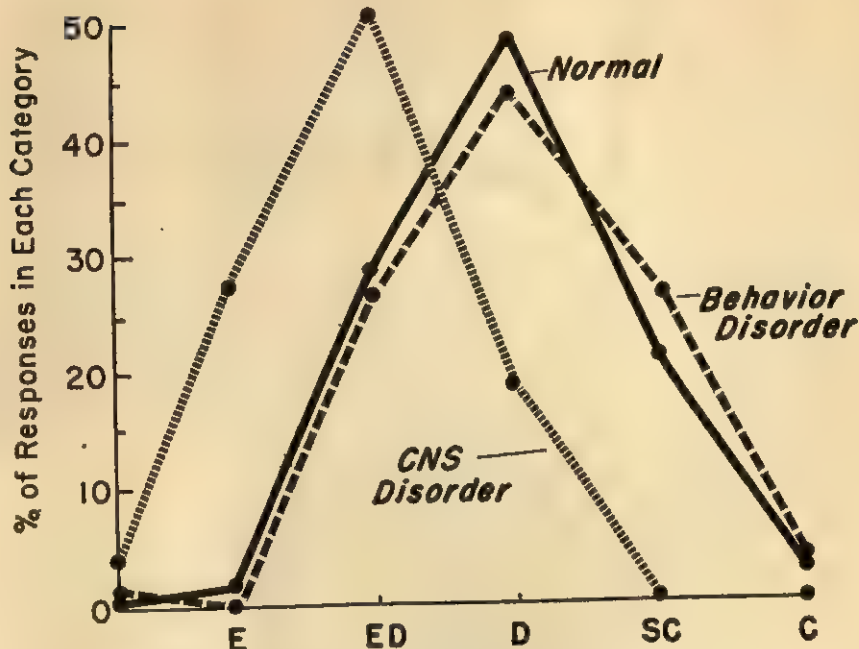


FIGURE 4  
PER CENT OF RESPONSES IN EACH CONTINUUM CATEGORY, PART III

TABLE 3  
GROUP COMPARISONS OF CONTINUUM SCORES

Group	N	Mean	SD	t
Normal	373	31.49	4.63	.59
Behavior Disorder	30	30.97	4.98	
Normal	373	31.49	4.63	11.51*
CNS Disorder	32	21.22	6.70	
Behavior Disorder	30	30.97	4.98	6.36*
CNS Disorder	32	21.22	6.70	

\* Significant at .001 level (two-tailed test).

## 2. Sex as a Factor

Exploration of sex as a factor in regard to the continuum did not reveal significant sex differences.



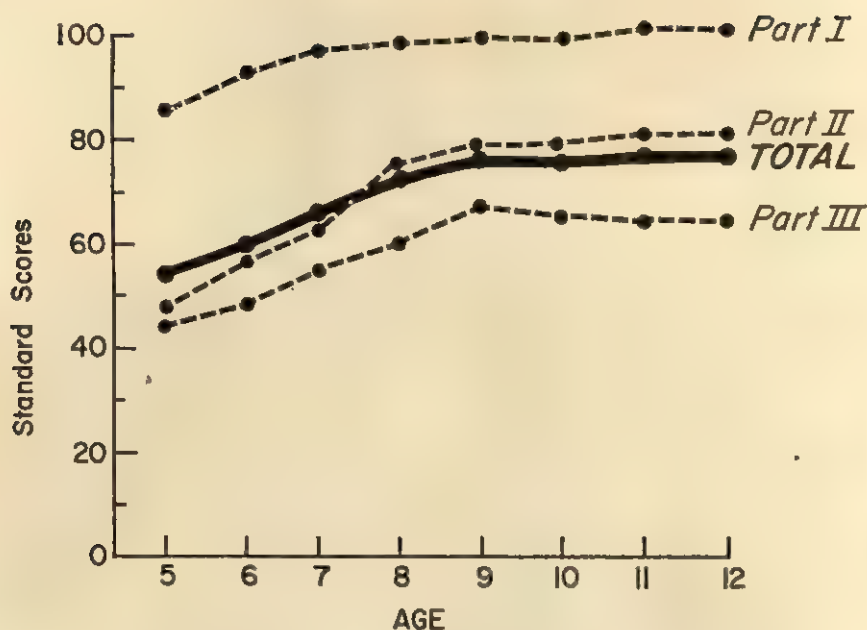


FIGURE 5  
AGE INCREMENTS OF CONTINUUM MEANS ON PARTS AND TOTAL OF BCT

TABLE 4  
CONTINUUM MEANS ACCORDING TO AGE

Age year	N	I		Test part II		III		Total	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
5	26	8.50	1.67	4.81	1.30	11.08	3.43	24.38	4.05
6	52	9.27	1.23	5.87	1.71	11.96	2.79	27.10	3.89
7	55	9.64	.79	6.27	1.88	13.62	2.95	29.53	4.09
8	53	9.79	.65	7.26	1.72	15.11	2.75	32.17	3.24
9	43	9.86	.51	7.63	1.67	16.84	2.17	34.33	3.06
10	45	9.91	.41	7.71	1.45	16.22	2.73	33.84	3.27
11	58	9.97	.26	8.09	1.18	16.09	2.37	34.14	2.71
12	41	10.00	—	8.07	1.39	15.93	1.85	34.00	2.19

TABLE 5  
CORRELATION OF CONTINUUM WITH AGE  
(Normal Group ( $N = 373$ ))

Continuum	$r^*$
Part I	.38
Part II	.51
Part III	.49
Total continuum	.62

\* With an  $N$  of 300 a correlation of .19 is significant at the .001 level.

#### D. DISCUSSION

The continuum was constructed on a theoretical basis and it was not known how the subjects would be distributed over the categories of the continuum. The subjects in the normal and the behavior-disorder groups fell in a statistically normal distribution over the continuum. These results are interpreted to mean that most of the children in these two groups were able to respond to the essential stimuli and to account for their being. Only a few subjects in these groups were able to attain the predictive categories. Thus, those who do fall at the upper end of the continuum should be considered superior individuals according to the theoretical definition of effective environmental interaction.

The investigators are not sure how to fit this finding into their theoretical approach. Since the prediction of response would seem to be a basic and fundamental type of behavior, more individuals should be capable of this response. A number of possibilities may be mentioned. For example, perhaps such failures are representative of problems in our training and in the education of children in our culture. Another possibility is that children predict responses only when motivation is greater as, for example, under threat or because of impelling forces due to specific forms of deprivation.

Although the investigators believe there is little relevance between the concepts of the continuum and those on which *IQ* is based, significant correlations were obtained between *IQ* and the continuum scores. These correlations are as follows: normal group, .22 ( $p < .01$ ); behavior disorder, .39 ( $p < .05$ ); CNS disorder, .39 ( $p < .05$ ).

#### E. SUMMARY

This study reveals how normal children and those with some form of psychopathology fall on a theoretical continuum of behavior. Children with CNS disorder tend to be aware only of the stimuli presented and fail to move to higher categories reflecting more complex behavior. In other words these children do not adequately use experience nor do they adequately predict future events. This study is a step toward defining the significance, meaning, and relevance of CNS disorder for the individual in his interaction with the environment. Children with behavior disorder, however, do not seem to be any more impaired in their ability to reason about external stimuli than do normal children. This finding suggests that behavior disorder is usually defined in terms of nonconformity to social expectations. In actuality there is little evidence to show that children so diagnosed behave any less effectively than do those considered to be normal.

## REFERENCES

1. ALEXANDER, T. The influence of central nervous system and behavior disorder upon complexity of response. *Amer. Psychologist*, 1961, **16**, 351.
2. ALEXANDER, T., & ALPERN, J. Psychological and physiological pathology of intellection. *Child Devel.*, 1960, **31**, 237-238.

*Department of Pediatrics*  
*University Hospitals*  
*State University of Iowa*  
*Iowa City, Iowa*

●



## CHANGES IN THE PHENOMENAL FIELD AS A RESULT OF HUMAN RELATIONS TRAINING\*<sup>1</sup>

*Department of Student Life, The City College of The City University of New York*

SUZANNE M. GASSNER, JEROME GOLD, AND ALVIN M. SNADOWSKY

### A. INTRODUCTION

Fundamental to the phenomenological approach to the understanding of human personality is the differentiation between the phenomenal self (including the self-concept) and the general phenomenal field (7). A relatively high degree of congruence between an individual's self-concept and his ideal self is considered to be an indication of good adjustment (6, pp. 55-75; 8; 13). Moreover, it has been found that an individual's acceptance of himself is positively and significantly correlated with his acceptance of others (2, 11, 12, 14).

Recently, a number of experiments reported that human-relations training leads to changes in the phenomenal self. Burke and Bennis (5), using an adaptation of Osgood's Semantic Differential to test the changes brought about by a three-week human-relations training program, reported a significant reduction in the discrepancy between the participants' self-image and ideal image. Grater (9) arrived at similar results using the Bills' Index of Adjustment and Values to test the effectiveness of an 11-week leadership training program for 30 college students. However, the hypothesized decrease in the discrepancy between the average other person and the ideal self was not found to be significant, though the results were in the predicted direction.

The City College of New York conducts a three-day human-relations training program (HRP) based on the design pioneered by Bradford, Lippitt, and Gibb (4). The aims of the program are to help the participants gain an understanding of behavior (both their own and that of others) and to provide an atmosphere conducive to change. Initial evaluation of the HRP was based on the participants' responses to a postmeeting reaction form. Reactions to the program were overwhelmingly enthusiastic, with 99 per cent of the students reporting that they had gained new skills and insights. Staff trainers have also

---

\* Received in the Editorial Office on February 10, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> We wish to thank the City College Fund for the grant-in-aid which supported the publication of this research.

reported that program participants became progressively more effective group members during the course of the HRP.

Based on these preliminary observations, it appears that The City College human-relations program, though considerably shorter in duration than those programs mentioned above, might nevertheless produce similar fundamental changes in the phenomenal self. It is therefore hypothesized that participation in the HRP causes an increase in the similarity between (a) an individual's self-concept and ideal-self concept and (b) an individual's self-concept and image of the average other.

The possibility that changes occur in areas of the participant's phenomenal field other than in his phenomenal self also seemed worth investigating. Bass (1) studied the kinds of changes that occur in the general phenomenal field as a result of training programs. He found that human-relations training leads to an increase in participants' sensitivity and understanding of interpersonal relations. Other kinds of attitude changes may be produced as a result of such programs. An integral part of the HRP is the demonstration and practice of skills required for effective democratic leadership. It is therefore expected that a significant increase in an individual's understanding of democratic leadership functions, and changes in his attitudes concerning such functions, occur as a result of participation in the HRP.

The first two experiments of this investigation, which parallel those reported in the literature, used a modified form of the Bills' Index of Adjustment and Values and a graphic rating scale to measure the changes that occurred in the divergence between the self-concept, the ideal-self concept, and the image of the other, for participants in the HRP. In the third experiment, attitudes toward democratic leadership functions, and understanding of them, were measured. However, unlike several of the research efforts referred to above, the instruments used were also administered to a control group.

## B. EXPERIMENT 1

### 1. *Method*

*a. Subjects.* Ss were 67 City College undergraduates. Forty-six were student trainees who attended the sixth semiannual HRP, the Experimental (EI) group. Twenty-one were students who submitted applications for the forthcoming program, the Control (C1) group.<sup>4</sup>

*b. Procedure.* One day prior to the onset of the HRP and two days after the completion of the Workshop, a modified form of the Bills' Index of Adjustment and Values (3) was administered individually to each subject. Ss were told that they were participating in a study to obtain information about

the beliefs and attitudes that C.C.N.Y. students have about themselves and other people. Using a series of 50 adjectives, *Ss* rated themselves on the following statements: (a) "This is characteristic of me," (b) "I would like this to be characteristic of me," and (c) "Most C.C.N.Y. students my age would like this to be characteristic of them." Each statement headed a separate page followed by the series of adjectives.

*c. Scoring.* The absolute discrepancies between the various statements on each of the adjective ratings were calculated for each *S*. The geometric mean was computed to provide a *D* score for each *S* for both administrations. Nonparametric statistics were used in the analysis of the data, because the distribution of the *D* values is not known.

## 2. Results

Results of the Wilcoxon matched-pairs signed-ranks test (Table 1) show that there was a significant increase in similarity between self-perceptions and ideal-self perceptions in both the E1 and C1 groups. In addition, both groups

TABLE 1  
INCREASE IN SIMILARITY BETWEEN SELF-IDEAL AND SELF-OTHER PERCEPTIONS  
SIXTH HRP—MODIFIED BILLS' INDEX OF ADJUSTMENT AND VALUES  
(Wilcoxon matched-pairs signed-ranks test with *z* score transformations)

Population	<i>T</i>	<i>N</i>	<i>z</i>	<i>p</i>
Self-ideal				
Experimental	242.5	46	3.26	.0006*
Control	18.0	21	3.40	.0006**
Self-other				
Experimental	87.0	46	4.96	<.0001*
Control	15.0	21	3.48	.0004**

\* One-tailed test.

\*\* Two-tailed test.

showed a significant increase in similarity between self-perceptions and average-other perceptions.

However, when the changes in Self-Ideal *D* scores (from the first to the second administration of the test) were examined, it was found that E1 group changes did not differ significantly from C1 group changes. Similarly, Self-Other *D* scores did not change in a significantly different fashion for the two groups (Table 2).



TABLE 2  
TESTS OF THE DIFFERENCES BETWEEN *D* SCORES FOR THE EXPERIMENTAL  
AND CONTROL GROUPS

Test	Experiment 1 ( <i>N</i> = 67)		Experiment 2 ( <i>N</i> = 72)	
	Self-ideal	Self-other	Self-ideal	Self-other
Rank test ( <i>z</i> )	.04	1.57	.44	1.59
Sign test ( $\chi^2$ )	.06	.57	.06	3.60

## C. EXPERIMENT 2

### 1. Method

*a. Subjects.* *Ss* were 72 City College undergraduates. Forty-five were student trainees who attended the seventh HRP (E2 group) and 27 were students enrolled in psychology courses at the college (C2 group).

*b. Procedure.* In the E2 group, a modified form of the Burke and Bennis graphic-rating-scale series was administered to the trainees at both the introductory session and the closing session of the HRP.

*Ss* were told that they were participating in a study of word meanings and that "the object of the study is to find out how you like to describe yourself and others."

Three concepts were rated against a uniform series of bipolar traits that were selected for their relevance to the attitudes being measured. The concepts were (*a*) the way I actually am in a City College group, (*b*) the way I would like to be in a City College group, and (*c*) the way other City College students generally act in a group. Each concept headed a separate page and was followed by the series of 18 scales.

The instructions given for the second administration were "to take into account your experiences during the past three days. You may feel free to change your original responses or to respond as you did initially. In any case your answers should be an accurate description of yourself and other people as you see them."

Control group *Ss* completed the graphic rating scale twice, with a three-day time interval, during regularly scheduled classroom sessions. The instructions were identical to those used for the E2 group.

*c. Scoring.* *D* scores were computed using the procedure described in Experiment 1.

### 2. Results

The results of the Wilcoxon matched-pairs signed-ranks test (Table 3) indicate that in the E2 and C2 groups there was a significant increase in simi-

TABLE 3  
INCREASE IN SIMILARITY BETWEEN SELF-IDEAL AND SELF-OTHER PERCEPTIONS  
SEVENTH HRP—GRAPHIC RATING SCALE  
(Wilcoxon matched-pairs signed-ranks test with  $z$  score transformations)

Population	$T$	$N$	$z$	$p$
Self-ideal				
Experimental	405.5	45	1.26	.10*
Control	156.0	27	.79	.43**
Self-other				
Experimental	232.0	45	3.21	.0007*
Control	105.0	27	2.02	.041**

\* One-tailed test.

\*\* Two-tailed test.

larity between self-perceptions and average-other perceptions. However, unlike the findings of Experiment 1, there was no significant increase in similarity between the self-concept and the ideal-self concept in either the E2 or C2 groups.

Using both the rank test for two independent samples and the sign test (Table 2), it was found that when the changes in the E2 and C2 groups were compared, the differences were not statistically significant.

#### D. EXPERIMENT 3

##### 1. Method

*a. Subjects.* Ss were 122 City College undergraduates. Ninety-four were student trainees who attended the seventh and eighth HRP (E3 groups) and 28 were students enrolled in psychology courses at The City College (C3 group).

*b. Procedure.* At the introductory sessions of the seventh and eighth HRP, a "Democratic Leadership Attitude Scale" was administered to a group of 43 and 51 trainees respectively. The 23 items of the scale were selected from the Ideology Questionnaire<sup>2</sup> by a panel of judges (staff trainers). Only questions that measured attitudes toward and understanding of democratic leadership were chosen. Ss were asked to rate the statements on a five-point scale (from strongly disagree to strongly agree). They were told that the statements referred to opinions regarding a number of leadership issues about which some people agreed and others disagreed. The second administration was conducted during the final evaluation session. Ss were

<sup>2</sup> Adapted from an Ideology Questionnaire developed by David Jenkins and others at Bethel National Training Laboratories in Group Development.

instructed to "feel free to change your original responses or respond as you did initially."

Control group Ss completed the first administration of the Democratic Leadership Attitude Scale during a regularly scheduled classroom session. Two days later Ss were again asked to complete the forms.

*c. Scoring.* Using the judges' responses as a criterion, each of the 23 items was assigned a correct answer that was either "agree" or "disagree." Subjects received points as follows: (a) If the correct answer was "agree," an answer of "agree or strongly agree" received +1; an answer of "disagree or strongly disagree" received -1; and an answer of "undecided" equalled zero. (b) If the correct answer was "disagree," the scoring was in reverse to the above. The maximum score was 23; the minimum, -23.

## 2. Results

Whereas both E3 and C3 groups were initially equal in their attitudes toward and understanding of democratic leadership concepts, the difference between the groups after the HRP was statistically significant (Table 4).

TABLE 4  
TESTS OF THE DIFFERENCES BETWEEN MEANS ON THE DEMOCRATIC  
LEADERSHIP ATTITUDE TEST

Group	N	Pretest	Posttest	t
Seventh HRP				
Experimental	43	12.7	16.4	5.32**
Control	28	12.6	12.3	.52
t		.098	3.94**	
Eighth HRP				
Experimental	51	12.5	15.8	6.75**
Control	28	12.6	12.3	.52
t		.096	3.27*	

\* Significant at .01 level.

\*\* Significant at .001 level.

This difference was caused by the increase in understanding of democratic leadership concepts that occurred in the experimental group.

## E. DISCUSSION

HRP trainees shifted their perceptions of themselves and others in the predicted directions. This aspect of the results is similar to the findings that Grater, and Burke and Bennis report. However, unlike the experiments they report, when a control group was used, it was found that similar changes in



the phenomenal-self attitudes could be demonstrated, and the extent of change between E and C groups was shown to be statistically insignificant. The control group adds an important dimension to this research because the information it provides forces a rejection of the initial conclusion that HRP participation causes change in the phenomenal self. If a control group had been used in previous research, the conclusions may have been modified or reversed and personality structure may have been found to be more stable than the earlier experiments have indicated.

Another possible interpretation of the results of earlier studies is that methodological inadequacies of self-rating scales counteracted attempts to measure changes in the phenomenal self. Wylie (16) suggests that a subject may try to present himself as having attitudes other than those that are actually true of himself, and that he may not be willing to reveal certain information about himself. A further consideration is the effect of familiarity on future responses to scale words. Taylor (15) found a marked increase in the similarity between the self-image and the ideal-self image on repeated tests conducted over a short time interval. Finally, Jouard and Lasakow (10) have pointed out that subjects will describe their attitudes and opinions much more openly than they otherwise will, when the things described do not relate directly to the subjects' own perception of their personality; such studies do not cause as much defensiveness on the part of the participants and thus contribute less to distortions in their responses. These interpretations suggest that conclusions pertaining to changes in the phenomenal self that occur over a short period of time should be re-examined.

Human-relations training seems to stimulate growth and understanding in areas not directly related to the phenomenal self. Because highly significant changes in the participants' understanding of and attitudes toward democratic leadership were found to occur during the training experience, while no such changes occurred for members of the control group, it was concluded that the HRP is effective in teaching certain leadership principles and in developing particular attitudes toward democratic methods. This study and the one conducted by Bass (1) point to other dimensions of the phenomenal field which have been demonstrated to change over a short time period. It might be worthwhile, therefore, to investigate further the kinds of changes in concepts, other than those pertaining to the phenomenal self, that may be the outcome of human-relations training.

#### F. SUMMARY

Recent research claims to demonstrate that changes in the phenomenal self occur as a result of human-relations training. Participants in a three-day

human-relations program were tested on an adjective checklist and a graphic rating scale which were used as indices of change in the phenomenal self. In general, predictions that a reduction in the trainees' discrepancy scores between the self-concept and both the ideal self and the image of the other were confirmed. Similar changes were observed to occur in the control group, but the differences between the results for the experimental and the control groups were not significant. However, when a measure of attitudes toward and understanding of democratic leadership functions was administered to both training participants and a control group, a highly significant change was found to occur for the experimental group, while no such change was observed in the control group. This information suggests that (a) personality structure may be more stable than the reports of recent experiments indicate, (b) methodological inadequacies of self-rating scales may interfere with attempts to measure changes in the phenomenal self, and (c) the measurement of attitudes other than those relating to the self is a fruitful area for further investigation in human-relations programs.

## REFERENCES

1. BASS, B. Reactions to Twelve Angry Men as a measure of sensitivity training. *J. Appl. Psychol.*, 1962, **46**, 120-124.
2. BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. Abn. & Soc. Psychol.*, 1952, **47**, 778-782.
3. BILLS, R. E. Manual for the Index of Adjustment and Values. Auburn: Alabama Polytechnic Inst., 1958.
4. BRADFORD, L. P., LIPPITT, G. L., & GIBB, J. R. Human relations training in three days. *Adult Leadership*, 1956, **4** (10), 11-26.
5. BURKE, R. L., & BENNIS, W. G. Changes in perception of self and others during human relations training. *Hum. Relat.*, 1961, **14**, 165-182.
6. BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counselling. In Rogers, C., & Dymond, R. (Eds.), *Psychotherapy and Personality Change*. Chicago, Ill.: Univ. Chicago Press, 1954.
7. COMBS, A. W., & SNYGG, D. Individual Behavior. A Perceptual Approach to Behavior (2nd ed.) New York: Harper, 1959.
8. COWAN, E. L., HEITZLER, F., & AXELROD, H. S. Self-concept conflict indicators and learning. *J. Abn. & Soc. Psychol.*, 1955, **51**, 242-245.
9. GRATER, M. Changes in self and other attitudes in a leadership training group. *Person. & Guid. J.*, 1959, **37**, 493-496.
10. JOUARD, S. M., & LASAKOW, P. Some factors in self-disclosure. *J. Abn. & Soc. Psychol.*, 1958, **56**, 91-98.
11. MCINTYRE, C. J. Acceptance by others and its relation to acceptance of self and others. *J. Abn. & Soc. Psychol.*, 1952, **47**, 624-625.
12. OMWAKE, K. The relation between acceptance of self and acceptance of others shown by three personality inventories. *J. Consult. Psychol.*, 1954, **18**, 443-446.
13. ROGERS, C. R. Client-Centered Therapy. Boston: Houghton Mifflin, 1951.

14. SHEERER, E. T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counselling cases. *J. Consult. Psychol.*, 1949, **13**, 169-175.
15. TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. Consult. Psychol.*, 1955, **19**, 205-209.
16. WYLIE, R. C. *The Self Concept: Critical Survey of Pertinent Research Literature*. Lincoln, Nebr.: Univ. Nebraska Press, 1960.

*Department of Student Life*  
*The City College*  
*139 Street and Convent Avenue*  
*New York 31, New York*



## ANALYSIS OF THE "GOOD ERRORS" MADE BY CLINICAL PSYCHOLOGISTS IN DIAGNOSTIC JUDGMENT\*

*Department of Psychology, Los Angeles State College*

---

JOSEPH G. PHELAN

---

### A. INTRODUCTION

In this study, an attempt is made to analyze the incidence of correct identifications of a person through test data and to investigate the relationship between degree of accuracy in judging and the *rationale* or kinds of reasons given for clinical judgments as made. Estes (10) found, when psychiatric social workers were asked to make judgments about personality characteristics, that some social workers demonstrate much greater ability to make use of cues than others. Existence or degree of this skill seemed not to be related to age, length of service, sibling status, or to whether or not the judge had been analyzed. Horn (12), working with Murray (16), used an analysis of variance technique to compare the relative importance of the ability of the judge and the aspect of personality measured. His judges differed greatly in ability even though similarly trained. Horn felt that, in presentation of the data and in teaching of the clinical process, it is important to single out or emphasize those aspects of personality that cause difficulty in diagnosis and to recognize the relative importance and weight to be attached to various kinds of clinical evidence.

A plan for isolating diagnostic skill (borrowed from the operational methods used in other disciplines) is a matching technique that requires the identification of a variety of data as belonging to one individual. Insight into the diagnostic process is gained by systematic analysis of the reasons given for matching. An analysis of errors constitutes a nonstatistical method of evaluating randomness of reasons given.

By Vernon's definition (22), "good errors" occur when the similarities between subjects outweigh the differences, so that, in many cases, test results for an apparently similar subject are substituted for those of the right subject. Because "good errors" can be evaluated statistically, a study of them affords insight into the ways in which the test results of apparently widely different persons can seem similar to an analyzing judge.

---

\* Received in the Editorial Office on February 10, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

## B. METHOD

The task of matching 16 units or documents was presented to 20 trained and experienced clinical psychologists having at least two years of clinical experience beyond the Ph.D. degree. To maximize the difficulty of the task and to keep the number of units to be matched within protocol limits, the documents were presented in four arrays with unequal matchings. In each array the four tests were representative of the same six individuals as follows:

### 1. *Array A*

Four standardized dictated Autobiographies, based on Murray's Form for Autobiography (16), were used.

### 2. *Array B*

Four complete TAT protocols were used.

### 3. *Array C*

Four complete Rorschach test results—with scoring and location charts—and four Holsopple-Maile Sentence Completion Test responses were used.

### 4. *Array D*

Four sets of responses to Thurstone's Primary Mental Abilities Tests, the Kuder Preference Record, and the Guilford-Zimmerman Temperament Tests were used.

## C. SUBJECTS

The subjects differed widely in diagnosis, age, and socioeconomic background. Vernon (23) stresses the influence on validity-matching of homogeneity and suggests as wide a diversity of materials as is practicable. The six subjects were as follows:

1. A 34-year-old male convict—college graduate—convicted for several offenses such as molesting young girls (probably schizophrenic; diagnosed by prison psychiatrist as psychopath-sexual deviate).

2. A 23-year-old male mental-clinic outpatient, unmarried (denied any hetero- or homosexual experience; diagnosis—anxiety state).

3. A 28-year-old female research scientist living successfully in the community (no psychiatric diagnosis or treatment).

4. A 16-year-old male high-school student under psychiatric treatment for anxiety and homosexual panic.

5. A 36-year-old female probationer convicted of prostitution (no diagnosis—bland, immature, without foresight or concern).

6. A 36-year-old male business executive (no diagnosis—not in need of treatment).

#### D. THE MATCHING TASK

Table 1 shows the scheme of presentation of the matching task. All materials were edited to eliminate extrinsic clues that might make matching

TABLE 1  
THE SCHEME OF PRESENTATION OF THE TASK (IN EACH ARRAY: FOUR UNITS,  
REPRESENTATIVE OF THE SIX SUBJECTS)

Type of record	Convict (male)	Mental- clinic out- patient (male)	Research scientist (female)	High- school student (male)	Pro- bationer (female)	Business executive (male)
Autobiography	X	X	X		X	
TAT		X	X			X
Rorschach and Sentence Completion	X	X		X	X	
Objective Battery (Primary Mental Abilities, Kuder Preference Record, Guilford-Zimmerman Temperament Tests)	X	X			X	X

possible without an evaluation of data. Whenever possible, topical references, turns of speech, or mannerisms were excised from the protocols. The instructions were as follows:

Appended hereto are 16 documents which are the results of standardized biographical interviews, objective tests, and projective techniques administered to six individuals. You are asked to indicate to which of the six individuals any two or more of these documents could be attributed, so that the documents which you assigned to one person could only have been produced by one person. Each protocol has identifying call letters. Indicate on the attached form your matching of two, three, or four documents as belonging to one person. Indicate whether you feel quite sure, comparatively sure, or unsure of your choice by checking the appropriate space.

Each of the documents is identified by a pair of letters. This pair is the designation to be used in describing the matching judgment. For ex-



ample, in filling out the accompanying form one might write: "Documents AB (biographical sketch), CD (Rorschach-Sentence Completion), EF (Thematic Apperception Test) seem to belong together," etc.

On the notation sheet that accompanied the instructions and test battery, each judge was asked to check one of three statements: "I felt very sure that my choice was correct," "relatively sure," "not at all sure."

### E. SCORING

A score of one for each correct match of any two documents was assigned. Judges were also given a score of one when they correctly indicated that one document had been omitted for any subject. Blanks in Table 1 correctly specified were given a score of one. The frequency of correct responses was compared with the frequency with which a given number of correct responses could be expected to occur by chance. Contingency coefficients were computed.

### F. RESULTS

Seven judges (A, E, H, I, M, N, R) performed at the  $p < .01$  level. Ten of the judges performed at a level of  $p < .05$  or better (Judges A, C, E, F, K, H, I, M, N, R). Ten others (B, D, G, J, L, O, P, Q, S, T) functioned at the level of chance expectation. In order to evaluate the performance of the judges as a group for significance, it is necessary to consider only the performance of those judges who completed the entire task. When the greatest number of correct arrangements of data can equal 18, the performance of the group, as a whole, exceeds chance expectation, with  $\chi^2 = 5.26$ ,  $p < .05$ .

#### 1. *Individual Differences in Matching Approach and Skill*

Using Holsopple and Phelan's computation of matching probabilities (8), we can consider mathematically the probability that a given number of successful matchings can occur by chance alone when one array is arranged against three other sets of data. Table 2 shows results for the judges who matched at a high level (11 out of 18 matches were correct) in contrast with judges who matched at a level no better than chance. Greatest success by most judges was achieved in matching projective tests to the Biography. Judges worked in different ways, depending on their training and preferences. Judge H worked well with all kinds of materials, but Judges C, K, L, and M (highly successful matchers in general) did much better in matching Rorschach to Biography than in matching Rorschach to TAT.

Table 2 presents the matching pattern of successful judges in contrast with

TABLE 2  
MATCHING PATTERN OF SUCCESSFUL VERSUS UNSUCCESSFUL JUDGES  
(JUDGES WITH 11 OR MORE CORRECT MATCHINGS)

Judge	Biography versus TAT	Biography versus Rorschach	TAT versus Rorschach	Judge	Biography versus TAT	Biography versus Rorschach	TAT versus Rorschach
	TAT	Rorschach	Rorschach		TAT	Rorschach	Rorschach
A	4	3	1	B	1	2	1
C	2	4	2	D	2	2	0
E	4	1	1	G	1	2	0
F	4	2	2	J	4	2	1
H	4	4	4	L	0	2	1
I	4	2	1	O	0	1	2
K	1	3	0	P	1	1	1
M	1	4	1	O	1	0	0
N	3	3	4	S	2	4	2
R	2	3	4	T	2	3	1

those of unsuccessful judges. All judges did much better when matching projective tests with Biography than when matching TAT and Rorschach, or in matching the objective test results with any of the projective test results.

Good judges exceed poor judges (*a*) by a 2-to-1 ratio when matching TAT and Rorschach, (*b*) by 20-to-1 when matching Biography and TAT, and (*c*) by 3-to-2 when matching Biography and Rorschach. Judges reported that they used Biography as the criterion against which to match other materials.

Table 3 shows the pattern of serial matching; i.e., the extent to which

TABLE 3  
PATTERN OF SERIAL MATCHING: EXTENT TO WHICH INDIVIDUAL JUDGES  
IDENTIFY SIX SUBJECTS IN SERIES WITH TESTS MATCHED AGAINST BIOGRAPHIES

Number of tests correctly matched against Biographies	Subject													
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	9	4	7	5	8	5	5	11	8	4	6	3	9	6
2	1		1		4	1	1	2	3	1	1		3	2
3	2							2						

individual judges were able to identify all six subjects in series, with tests matched against the biographies. Judges who did well in matching in pairs also did best in matching in series. Judges A, E, and H were particularly successful in using the approach of matching all tests against Biography.

## 2. Test Validity

By use of the  $\chi^2$  technique, it is possible to throw light on the validity of projective materials. When 14 judges complete the entire task of matching

one array against each other array, the performance of the judges as a group can be evaluated. Theoretically, in the matching of any array against any other array, we should expect that eight judges should match less than two subjects correctly and that six judges should match more than two. Our data on  $\chi^2$  values are presented in Table 4.

TABLE 4  
CHI-SQUARE VALUES WHEN EACH ARRAY IS MATCHED AGAINST  
EVERY OTHER ARRAY

Matching comparison	Less than two	Two or more	$\chi^2$	<i>p</i>
Biography versus Rorschach	2	12	12.6	.001
Biography versus TAT	5	9	3.15	.10
Biography versus Objective tests	5	9	3.15	.10
TAT versus Rorschach	10	4	1.4	.30
TAT versus Objective tests	6	8	1.4	.30
Rorschach versus Objective tests	5	9	3.15	.10

### 3. *Analysis of Reasons for Matchings Made by Judges*

Judges were asked to verbalize freely to reveal something about the cues underlying the diagnostic process. Only 13 of 20 judges gave reasons.

Eight of the judges who gave reasons were successful; five performed at the chance level. Because the reasons were incomplete and random, there is little point in statistical evaluation or in presenting them in terms of the accuracy of the corresponding judgments. Despite the fact that judges can make correct matchings at a significant level, their rationales for the judgments are not consistent. Because the clinician-judges did not meet or interview the subjects, the matching may be a result of successful detective work involving both clinical and nonclinical cues.

In general, comments of successful judges (A, F, H, M, R, S, T, U) were more numerous, lengthier, and more detailed than those of unsuccessful judges (B, G, K, L, N). Judges commented on any number of subjects from two to six.

## G. DISCUSSION

Several observations relative to the diagnosis of "openness" or distinguishability of the subjects who were analyzed are as follows:

Subjects were chosen with an eye to heterogeneity of sociocultural backgrounds. Therefore it was not possible in advance to determine whether judges would have difficulty working with one subject rather than with



another or whether, in any given comparison, subjects would seem more alike than different.

The subjects of this experiment (because of great effort in the preliminary stage of the work to build friendship and to enlist cooperation) tended to be open, talkative, and interesting. Several reported that they enjoyed taking the tests.

The subjects tended to be various and striking. Each subject seemed distinguishable from each other subject to the same degree—roughly. The female probationer and the older businessman seemed most easily distinguishable; the mental outpatient, least readily identified.

Biographies were not available for the high-school boy or for the older businessman and, apparently, it was not difficult to recognize that none of the biographies could match the tests for these individuals.

TAT records were withheld for the convict and for the female probationer. The omission of the TAT record for the convict caused 66 per cent of the judges to match the TAT for the college professor's wife to the convict's Biography.

### *1. Tendencies to Confusion*

Signs of extreme immaturity led judges to confuse the high-school boy with the mental-hygiene outpatient.

Objective tests of the bright, verbal, extrovert businessman were consistently identified with those of the intelligent college wife.

All tests for the repressed, dull-normal, impulsive clinic outpatient were assessed as those of the female probationer.

The Rorschach for the convict was frequently identified as that of the man who was known to have attended the mental-hygiene clinic.

Differences in the degree to which one or another personality seemed accessible to analysis seemed not great. Differences in the degree of success that individual judges attained on the overall task were as great as were differences in the degree of success of the group in working with different types of tests.

### *2. System*

Clinicians volunteered that they tended to employ a system in working with the data: (a) Some clinicians studied the Biography, matched the TAT to that, and then tried to visualize the individual. Other tests were then matched against this nucleus. (b) Other clinicians matched Biography and Rorschach and then tried to fit other tests into the pattern that they achieved by this means.

Clinicians were given freedom to use the procedure that seemed most natural to them (as based on their own familiarity with and use of the various tests).

The clinical approach seemed to consist of building a "structure" after the judge had identified the characteristics or a cluster of the characteristics of an individual. The common *modus operandi* seems to have been the conceptualization of an individual and a matching of all tests in terms of this picture.

It is conceivable that, in some cases, clinicians who obtain high scores, when their performance is evaluated in terms of matching in pairs of tests, tend not to be so successful in identifying a series of tests as belonging to one individual. Perhaps familiarity with the richness derived from one or two instruments makes the judge overlook relationships implied by other materials.

### 3. *A Study of the "Good Errors" Made by the Judges in Matching*

"Good errors," by Vernon's definition (23), occur when the similarities between subjects outweigh the differences, so that test results for a "wrong" subject are substituted for those of the right subject. These errors cannot be evaluated statistically, but a study of them affords insight into the ways in which the test results of apparently widely different persons can seem similar to an analyzing judge. Cases which can be described as "good errors" are listed in Table 5.

TABLE 5  
GOOD ERRORS

Tests	Number	Per cent
Biography (convict) with TAT (college professor's wife)	14	.666
Biography (clinic patient) with TAT (female probationer)	8	.381
Biography (convict) with Rorschach (clinic patient)	8	.381

A consistent trend appears in the misidentifications. Reasons given for mistakes provide details concerning the common categories which clinical workers have in mind when they classify materials. (See Tables 6, 7, and 8.)

The Rorschach record of the convict seems to be that of a psychotic subject, although there was no direct statement in the convict's Biography to the effect that he was psychotic. Unless his Biography is carefully inspected for internal cues, it might be assumed that any mental-clinic patient

TABLE 6  
CASE A: CONFUSION OF BIOGRAPHY OF THE CONVICT WITH TAT OF THE  
COLLEGE PROFESSOR'S WIFE

Reasons	Number	Per cent
1. Inability to compete with males, sexual inadequacy, taunted about something small	7	.124
2. Methodist religion	5	.088
3. Reaction to strict, rigid father; social pressure	4	.071
4. Reaction to both parents' pressure	3	.055
5. Reaction to mother, ties, pressure	3	.005
6. Exhibitionism, display of virility, publicity	5	.088
7. Vocabulary, reference to school, education, expressions	7	.124
8. Preoccupation with sex	4	.071
9. Voyeurism	2	.036
10. Sex play with sister, early sex prohibition, sex seen as hopeless, sexual deviate, poor control of sex, need for assault on children	7	.124
11. Homosexual guilt, compulsion to sex play due to homosexual panic	3	.055
12. Social withdrawal, failure	1	.018
13. Violence, sadism, aggression	3	.055
14. Depression	2	.036
Total	56	1.000

might produce such a record. It was attributed to the clinic patient by eight judges.

The "incorrect" reasons that clinicians gave to account for decisions in matching protocols that can be called "good errors" may be classified as in Table 9.

TABLE 7  
CASE B: CONFUSION OF BIOGRAPHY OF MENTAL CLINIC OUTPATIENT  
WITH TAT OF FEMALE PROBATIONER

Reasons	Number	Per cent
1. Attitudes female	3	.143
2. Vocabulary, vulgarism, education, low cultural level, narrow interests	5	.240
3. References to "movies"	3	.143
4. Immaturity, childishness, irresponsibility	3	.143
5. Represses guilt, lives for moment	3	.143
6. Guilt	2	.094
7. Sexual promiscuity	1	.047
8. Interest in dancing	1	.047
Total	21	1.000



TABLE 8  
CASE C: CONFUSION OF BIOGRAPHY OF MENTAL-CLINIC PATIENT WITH THE  
"PSYCHOTIC" RORSCHACH OF THE CONVICT

Reasons	Number	Per cent
1. Rorschach: feeling of split in two; Biography: "I feel like the two guys"	2	.071
2. Rorschach—Sentence Completion: references to sanity and insanity	1	.036
3. Sex dreams, fantasy	3	.107
4. Rorschach: Impulsiveness, acts at wedding (Biography)	1	.036
5. Rorschach: golf course; Biography: eye injury from golf club	1	.036
6. Rorschach: threatened disaster	2	.071
7. Rorschach: psychotic; Biography: hospitalization	3	.108
8. Obsessive-Compulsive	2	.071
9. Language peculiarities	1	.036
10. Fears of girl being raped	2	.071
11. Sex problems	5	.178
12. Sexual confusion	1	.036
13. Avoidance of women, sexual confusion, difficulties with women	4	.143
Total	28	1.000

Judges were frequently misled because of the following similarities between subjects which appeared to outweigh differences:

1. The female probationer and the mental-hygiene-clinic outpatient have similar intellectual levels, their cultural levels are similar, and they refer to the same movies and magazines. Both subjects are described as immature, impulsive, and nonintrospective.

2. The college professor's wife and the convict are both college educated, well read, and both speak well. Both have had some scientific training. Expression of their reactions to similar early religious training is parallel, as is reaction to parents. Although of opposite sex, both make many direct and indirect references towards striving for masculinity. Sex identification is similar, as is reaction to marriage partner.

3. The Biography of one subject referred to him as an outpatient at a mental-hygiene clinic. Many judges matched this Biography with the Rorschach and the TAT, which appeared to be that of a disturbed subject. The most seriously disturbed of the subjects was the convict, who had never been accurately diagnosed. His Biography gave other evidences of disturbance.

Analysis of reasons given for misidentifications indicates similarities in attitudinal and motivational patterns between subjects. Some frequent and consistent errors occur because of intrinsic similarities; others occur because

TABLE 9  
 "INCORRECT" REASONS THAT CLINICIANS GAVE TO ACCOUNT FOR DECISIONS  
 IN MATCHING PROTOCOLS THAT CAN BE CALLED "GOOD ERRORS"

Reason	A	B	C
1. Mention of an attribute or attitude that the two subjects have in common	1 3 7 11 12	1 2 4 5 6	3 4 8 11 12 13
2. Reason that seems based on an inadequate consideration of all phases and all inconsistencies in the data	1 4 8 9 10	1 7 8	1 2 7 8 9
3. Irrelevant or surface similarity	2	3	5
4. Details that do not appear in more than one record	4 5 6 8 9 10	7 8	2 6 10

judges are quick to jump at superficial similarities and to ignore contradictions in the data.

#### H. SUMMARY

The purpose of this study is to throw light on the clinical psychologist's judgmental ability. In diagnosis, the data obtained from psychological examinations are interpreted by the examiner in such a way that the data become meaningful in terms of the particular individual examined. In a matching task of 16 documents, judges matched documents and gave reasons for their matching judgments. Judges as a group performed at a higher level than could have been expected by chance. Some judges as individuals also performed considerably above chance. Others did not. Performances were differential, but judges who were superior in matching with any two tests were superior in most. Judges were able to identify correctly tests in series. Without having been so instructed, judges used the Biography as the criterion against which to match objective and projective test results.

There exists a diagnostic or matching skill. Some judges exhibit it in greater degree than others. The matching test technique seems to indicate which judges possess such skill.

The most important of the extraneous factors that influence the matching performance of clinical-psychologist-judges is the homogeneity or diversity

of materials judged. Homogeneity is dependent on the distinctiveness or range of unlikenesses among the subjects whose modes of expression and reflections of personality characteristics make up the data. A consideration of the mismatches indicates that the subjects frequently had characteristics in common, when the data were closely analyzed, that were not immediately apparent in the first inspection; i.e., similarities in viewpoint, attitude, mood, and dynamics. "Good errors" occur when similarities between subjects outweigh differences, so that the test results for similar subjects are misidentified. Such errors were not numerous in this study and did not occur so often as to invalidate the matches made.

## REFERENCES

1. ALLPORT, G. W. *Personality, a Psychological Interpretation*. New York: Henry Holt, 1937.
2. ———. The use of personal documents in psychological science. *Bull.* 49, Soc. Science Research Council, N.Y., 1942. Pp. 1-210.
3. ALLPORT, G. W., & VERNON, P. E. *Studies in Expressive Movement*. New York: MacMillan, 1933. Pp. 1-269.
4. CHALLMAN, R. C. The clinical psychology program at Winter V. A. Hospital. *J. Clin. Psychol.*, 1947, **3**, 28-31.
5. CHAPMAN, D. W. The statistics of the method of correct matchings. *Amer. J. Psychol.*, 1934, **46**, 287-298.
6. ———. The generalized problem of matchings. *Ann. Math. Stat.*, 1935, **6**, 85-95.
7. ———. The scoring of matching tests with unequal series. *J. Educ. Psychol.*, 1936, **27**, 368-370.
8. COMBS, A. A method of analysis for the Thematic Apperception Test and autobiography. *J. Clin. Psychol.*, 1946, **2**, 167-174.
9. CRONBACH, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.*, 1949, **46**, 393-429.
10. ESTES, S. G. Judging personality from expressive behavior. *J. Abn. & Soc. Psychol.*, 1949, **44**, 329-344.
11. HOLSOPPLE, J. G., & PHELAN, J. G. The skills of clinicians in analysis of projective tests. *J. Clin. Psychol.*, 1954, **10**, 317-320.
12. HORN, D. An experimental study of the diagnostic process in the clinical investigation of personality. Unpublished Ph.D. thesis, Harvard University, Cambridge, 1943.
13. KELLY, E. L. Research on evaluation of clinical psychologists. *J. Consult. Psychol.*, 1949, **13**, 34-38.
14. KELLY, E. L., & FISKE, D. W. The prediction of success in the V. A. Training Program in clinical psychology. *Amer. Psychologist*, 1950, **5**, 395-406.
15. KRUGMAN, J. L. A clinical evaluation of the Rorschach with problem children. *Rorschach Res. Exch.*, 1943, **6**, 61-70.
16. MURRAY, H. A. *Explorations in Personality*. New York: Oxford Univ. Press, 1938. Pp. 245-248.
17. PHELAN, J. G. Subjective feelings of certainty of diagnostic judgment of clinical psychologists. *J. Clin. Psychol.*, 1960, **16**, 101-104.
18. POLANSKY, N. A. How shall a life history be written? *Charac. & Personal.*, 1941, **9**, 188-207.

19. TROOP, H. A comparative study by means of the Rorschach method of personality development in 20 pairs of identical twins. *Genet. Psychol. Monog.*, 1938, **20**, 461-536.
20. VALENTINE, C. W. The relative reliability of men and women in intuitive judgements of character. *Brit. J. Psychol.*, 1929, **19**, 213-328.
21. VERNON, P. E. Some characteristics of the good judge of personality. *J. Soc. Psychol.*, 1933, **4**, 57.
22. ———. The matching method applied to investigations of personality. *Psychol. Bull.*, 1933, **33**, 149-177.
23. ———. Evaluation of the matching test. *J. Educ. Psychol.*, 1936, **27**, 1-17.
24. WALLIN, P. The prediction of individual behavior from case studies. In P. Horst (Ed.), *The Prediction of Personal Adjustment*. New York: Social Science Research Council, 1943.
25. WOLF, R., & MURRAY, J. A. An experiment in judging personality. *J. of Psychol.*, 1936, **3**, 345-365.
26. ———. Judgements of personality. In H. A. Murray (Ed.), *Explorations in Personality*. New York: Oxford Univ. Press, 1938. Pp. 243-281.

*Department of Psychology*  
*Los Angeles State College*  
*5151 State College Drive*  
*Los Angeles 32, California*



LYSERGIC ACID DIETHYLAMIDE (LSD-25):  
XXXV. COMPARISON OF EFFECT ON  
SIAMESE FIGHTING FISH  
AND GOLDFISH\*<sup>1</sup>

*South Oaks Research Foundation, Inc., Amityville, New York*

---

H. H. GETTNER, A. ROLO, AND H. A. ABRAMSON

---

A. INTRODUCTION

Data on the effect of LSD-25 on the behavior of Siamese fighting fish (*Betta splendens*) has been published hitherto (1, 2). Less than 0.5 mcg per ml of LSD-25 in the outside liquid causes surfacing of the fish and relative immobility. Injection of 0.01 mcg LSD-25 produces similar effects in fish weighing about one gm. The Siamese fighting fish may therefore be used as a test animal for the bioassay of LSD-25 and other psychotomimetic drugs. Figure 1 illustrates older data on the effect of LSD-25 on *B. splendens*. These older data are given by the dotted lines. Newer data are represented by the heavy solid lines and will hereafter be used in our experiments as reference curves. Figure 2 employs the data of Figure 1 so that the surfacing behavior described in detail in the second reference are replotted as a function of minutes for the first 70 minutes of the experiment. Note that in the case of *B. splendens* with our former technique, the reaction to 0.5 mcg of LSD-25 in the outside liquid is markedly less than the reaction to 1.0 mcg of LSD-25 under the same conditions. In an attempt to find an inexpensive and hardier fish, we have used ordinary goldfish (*Carassius auratus*). These are readily purchased in pet shops. In addition, at the suggestion of Dr. Howard A. Loeb of the New York State Conservation Department, we have done a few preliminary experiments on minnows, kindly supplied by him.

B. EXPERIMENTAL

1. *Effect of LSD-25 on Goldfish*

The technique used with Siamese fighting fish was modified with goldfish because surfacing of the goldfish is not readily observed if the liquid surface is large. This was originally observed by Loeb (3) during his experiments

\* Received in the Editorial Office on February 18, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported in part by a grant from the Sport Fishing Institute, Washington, D. C.

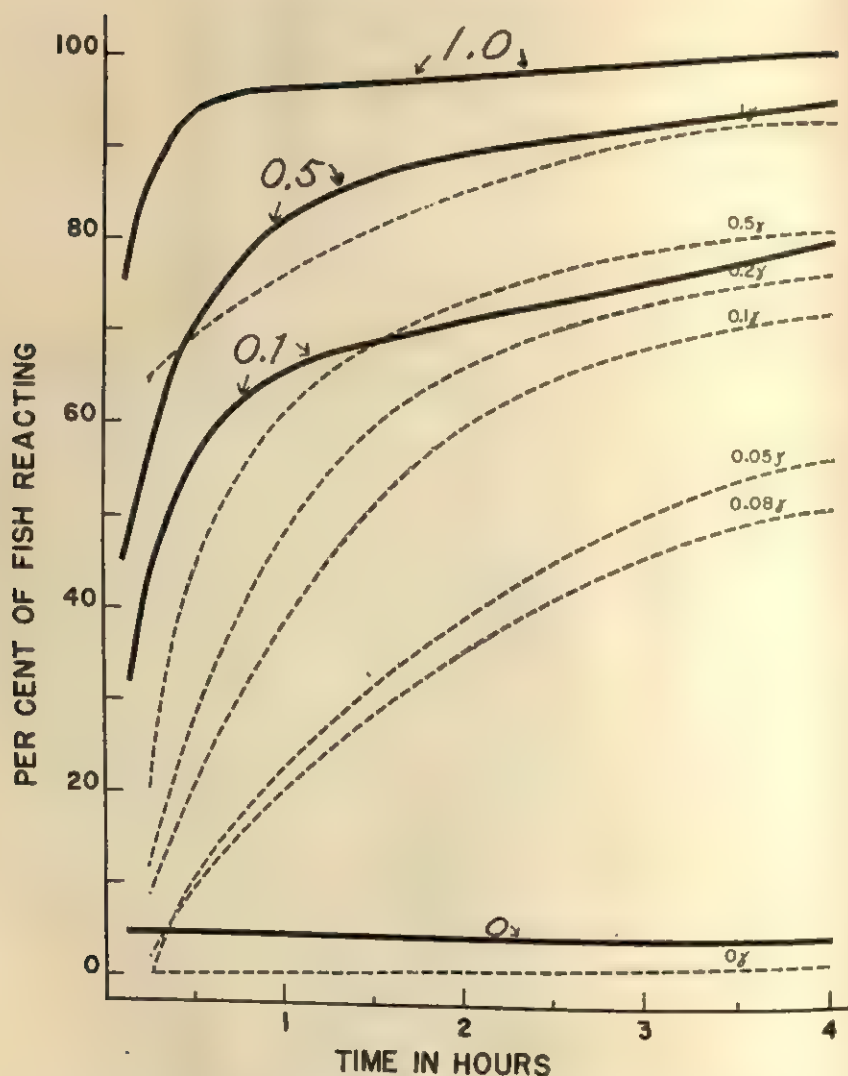
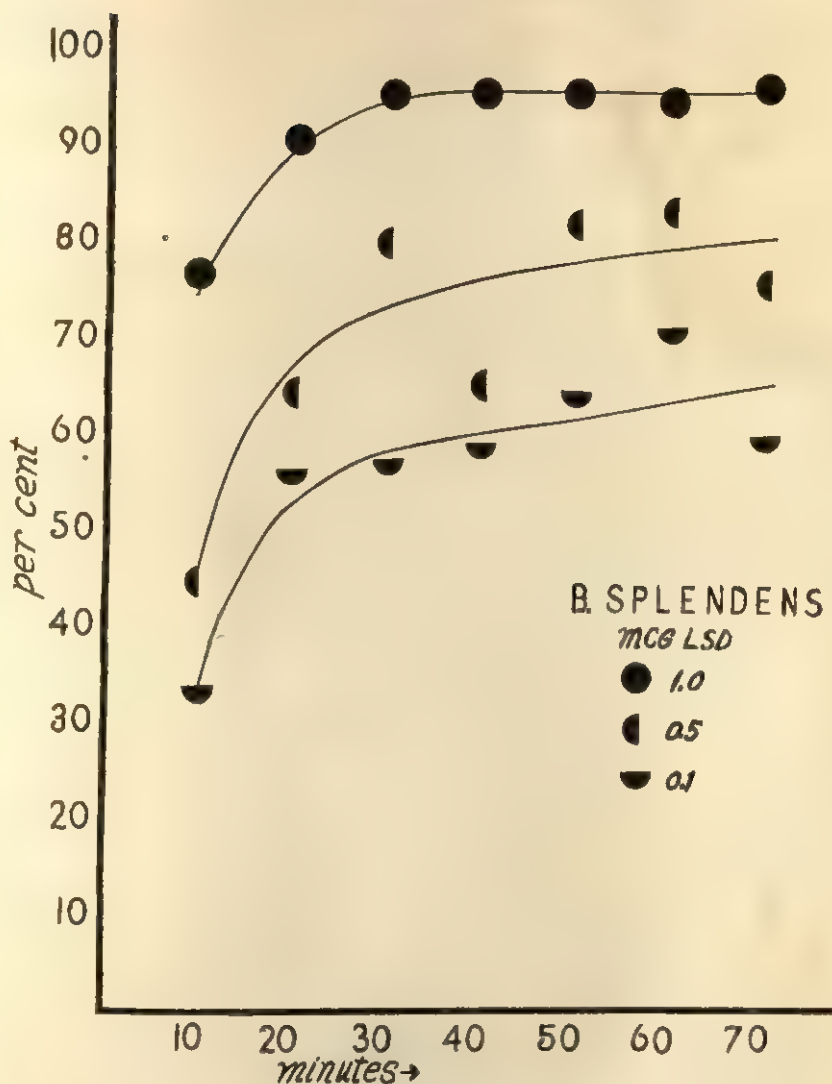


FIGURE 1  
PER CENT OF *B. Splendens* REACTING TO INJECTIONS OF LSD-25 AS  
A FUNCTION OF TIME IN HOURS

The dotted lines represent older data on the effect of LSD-25 on *B. splendens*. The heavy smooth curves labeled 0.1, 0.5, 1.0 are new standard values for mcg of LSD-25 in the outside liquid. These data were obtained from a paper previously published.



• FIGURE 2

PER CENT OF *B. Splendens* REACTING TO INJECTIONS OF LSD-25 AS  
A FUNCTION OF TIME IN MINUTES

In order to visualize the new data presented here on goldfish, data on surfacing behavior from Reference 2 are replotted as a function of minutes for the first 70 minutes of the experiment. Note in the case of *B. splendens* with our former technique, the reaction to 0.5 mcg of LSD-25 in the outside liquid is markedly less than the reaction to 1.0 mcg of LSD-25 under the same conditions.

with large carp. Loeb found that carp, which surfaced in a small tank, swam more freely if transferred to a pond. By indirection, this observation probably accounts for many failures to observe surfacing of Siamese fighting fish in large tanks.

Ten goldfish, one to one and one-half inches in length are placed in one-liter graduate cylinders (diameter = 6.4 cm), each of which contains one liter of distilled water. A slow steady stream of air is bubbled through each cylinder throughout the whole experiment. The water in the cylinders is brought to the same temperature as the room-temperature water in the tanks from which the fish are taken. After the fish have been in the cylinders from 15 minutes to one-half hour in order to adjust to the altered environment, the LSD-25 is added. The experiments to be reported here lasted one hour and 10 minutes.

Concentrations of 0.1, 0.2, 0.5 and 1.0 mcg LSD-25 per ml have been tested on the goldfish. The data are illustrated in Figure 3. In concentrations of 0.5 and 1.0 mcg, substantially all of the fish are swimming at the top of the cylinder—nose up, tail down—in 10 minutes. Due to the airstream, there is some swimming away from the top and returning. This pattern persists for the duration of the experiment. In the lower concentrations, there is more swimming away from the top, and the fish seem to "hang" at the top less than in the higher concentrations. In concentrations of 0.2 mcg LSD-25 per ml, it takes 20 to 30 minutes for 70 per cent or more of the fish to react. With 0.1 mcg LSD-25 per ml, it takes 30 minutes for 70 to 80 per cent of the goldfish to react. In distilled water, from 20 to 25 per cent of the fish occasionally swim at the top, but this per cent has been decreased by the waiting time at the beginning of the experiment. Figure 4 represents the average per cent of fish which surface when exposed to LSD-25 in the outside liquid, as plotted against time. Note that goldfish under experimental conditions are very sensitive to less than 1.0 mcg of LSD-25 in the outside liquid.

## 2. *Effect of LSD-25 on Minnows*

The technique with goldfish may not be used with minnows of the size studied. The graduate cylinders used for the goldfish proved unsuitable; the experimental fish and the controls died in this confining environment. Five minnows (two and one-half inches to four inches in length) were placed in three liters of distilled water in a four-liter beaker (diameter = 20 cm) with a rapid stream of air bubbling through the liquid during the entire experiment. The beakers are not filled, as the fish tend to jump out of filled vessels—even



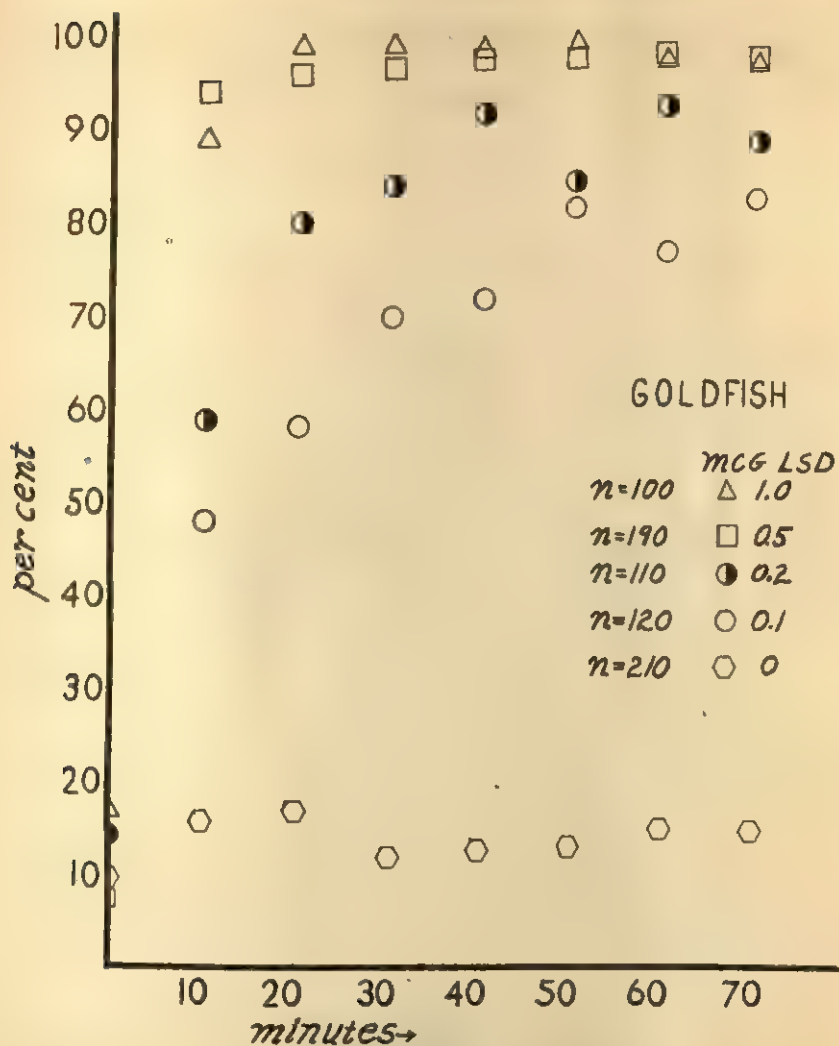


FIGURE 3  
PER CENT OF GOLDFISH REACTING TO INJECTIONS OF LSD-25  
AS A FUNCTION OF TIME IN MINUTES (RAW DATA)

The data on goldfish show that goldfish studied under conditions described in the text are at least as sensitive to LSD-25 in the outside liquid as Siamese fighting fish. Smoothed curves for these data are presented in Figure 4.

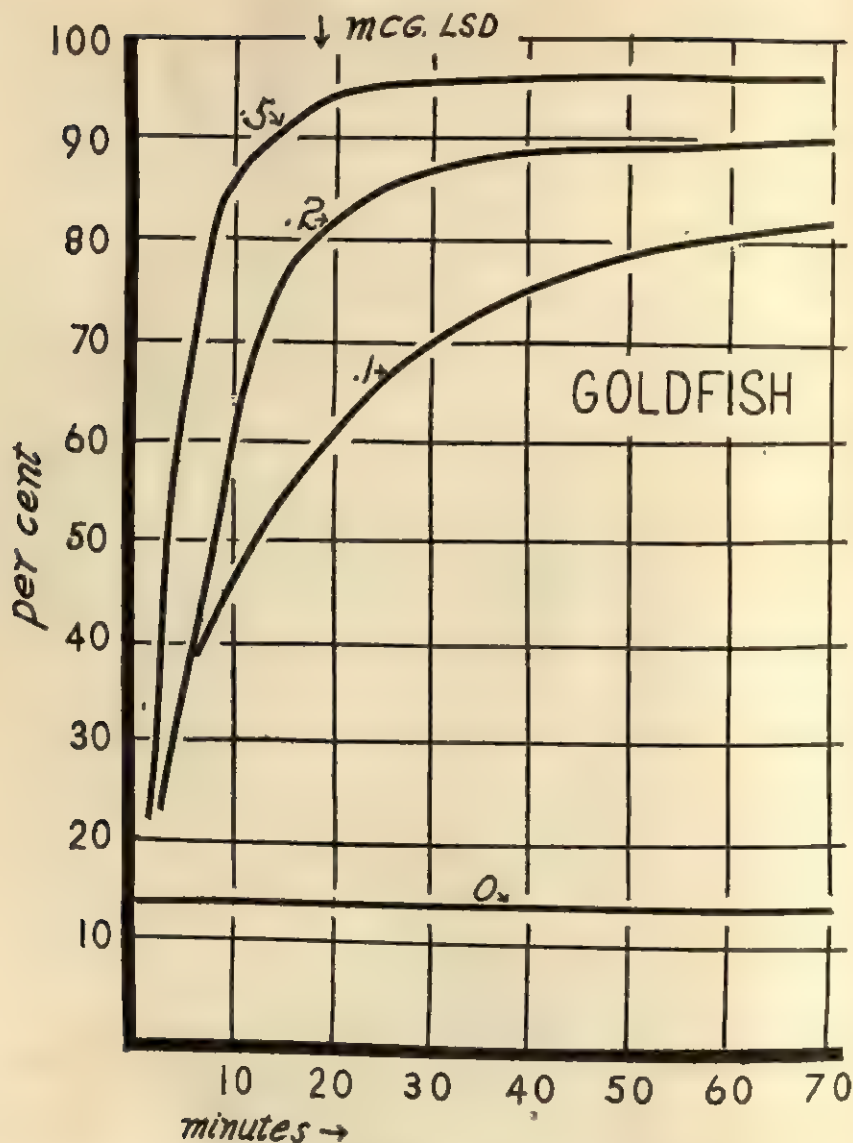


FIGURE 4  
PER CENT OF GOLDFISH REACTING TO INJECTIONS OF LSD-25  
AS A FUNCTION OF TIME IN MINUTES (SMOOTHED DATA)

The average per cent of fish surfacing when exposed to LSD-25 in the outside liquid is plotted against time. Note that goldfish, under the experimental conditions, are very sensitive to less than 1.0 mcg in the outside liquid.

large tanks. From the few experiments that we have done, we have found that minnows (like goldfish and Siamese fighting fish) all readily surface with concentrations of 0.5 mcg LSD-25 per ml; and with 0.2 mcg LSD-25 per ml, in 20 minutes 80 per cent will surface.

At the suggestion of Dr. Howard Loeb, we have run preliminary experiments with both goldfish and minnows using n-iso-propyl-d-lysergic acid amide maleate in concentrations of 1.0 mcg and 0.5 mcg per ml. This congener of LSD-25 causes surfacing in both goldfish and minnows.

### C. CONCLUSION

Readily available and hardy fish like goldfish and minnows have been found to be as sensitive to LSD-25 as the Siamese fighting fish which we have used in earlier experiments. Though the goldfish and the minnows do not exhibit all the effects shown by the Siamese fighting fish, the surfacing reaction—with or without the nose-up, tail-down position—is as reliable and as sensitive as the reactions of the Siamese fighting fish. Either the goldfish or the minnows will prove good test animals for derivatives and congeners of d-lysergic acid. It is important that the experimental technique be adapted to the surfacing behavior with special reference to the facilitation of optimal conditions for clustering at the surface. Standard curves are provided.

### REFERENCES

1. ABRAMSON, H. A., & EVANS, L. T. Lysergic acid diethylamide (LSD-25): II. Psychobiological effects on the Siamese fighting fish. *Science*, 1954, **120**, 990-991.
2. ABRAMSON, H. A., GETTNER, H. H., ROLO, A., & DEAN, G. Lysergic acid diethylamide (LSD-25): XXXIV. Comparison with effect of psilocybin on the Siamese fighting fish. *J. of Psychol.*, 1963, **56**, 363-374.
3. LOEB, H. A. Effect of lysergic acid diethylamide (LSD-25) on the surfacing behavior of laboratory carp. *N.Y. Fish & Game J.*, 1962, **9**, 127-132.

*South Oaks Research Foundation, Inc.*

*Amityville, New York*

## SPECIFIC INTERFERENCE IN RETROACTIVE INHIBITION\*<sup>1</sup>

*Department of Psychology, Yeshiva University*

---

JOHN CERASO

---

### A. INTRODUCTION

Interference theories of forgetting derive much of their factual support from studies that demonstrate that recall is a function of the similarity of the stimuli of original and interfering material. In an RI experiment, interference is thought to operate by either the transfer of the original responses to the interpolated learning, at which point the original responses are extinguished; or by the transfer of interpolated responses to the point of recall, at which point they can compete with original responses. Thus, interference theories are essentially based on transfer notions, and the facts concerning similarity and interference are congruent with our knowledge of transfer.

There is an issue, however, concerning transfer which has not been resolved. This issue revolves about the question as to what is the unit that is transferred. Starting from the Thorndikian notion of transfer of identical elements, many theorists have assumed that the unit of transfer is the specific stimulus-response pair. For example, at the point of recall, the stimulus presented will elicit all responses that have been learned to stimuli identical or similar to the stimulus presented. A somewhat different possibility is currently receiving attention (3, 4) which considers the unit of transfer as being the entire list. According to this position, the stimulus presented at recall will tend to elicit all responses from the interfering list. This type of theory, then, assumes that interference is general or nonspecific.

What is the evidence for nonspecific interference? Two lines of evidence seem most persuasive. One is the finding that general environmental similarity exerts an influence on interference (1, 2). The second is Runquist's finding (4) that recall of a given A-B pair is independent of the learning strength of the corresponding A-C pair for both RI and PI. These findings could be taken to mean that in some way the entire competing list disrupts the recall of the original list. Stimulus similarity could have the effect of making *lists*, and not specific *responses*, more likely to interfere with each other.

---

\* Received in the Editorial Office on February 19, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Experiment I is part of a dissertation submitted to the Graduate Faculty of the New School for Social Research. The remaining experiments were supported by a grant, G 19538, from the National Science Foundation.



The experiments to be reported have as their aim the further investigation of specific interference in RI. The technique used to study this problem was the mixed-list (ML) technique previously used in the study of transfer (7). The design of the basic experiment to be reported is presented in Table 1, with only four OL and four IL pairs indicated. Note that each pair is des-

TABLE 1  
THE MIXED-LIST DESIGN

Pairs	Original Learning	Interpolated Learning
Repeated pairs	$A_1-B_1$	$A_1-B_5$
	$A_2-B_2$	$A_2-B_6$
Nonrepeated pairs	$A_3-B_3$	$A_5-B_7$
	$A_4-B_4$	$A_6-B_8$

ignated by the symbols A-B, while subscripts serve to identify which stimuli in the first list are repeated or not repeated in the second list. Since no responses are repeated in the second list, it should be realized that the interpolated pairs are either A-C or C-D with respect to the first list. The basic idea of the experiment is to compare the recall of the repeated (R) pairs with the recall of the nonrepeated (NR) pairs. If interference is essentially by the entire second list, then obviously R and NR pairs are interfered with by the same second list, and therefore recall for these pairs should, according to a nonspecific interference theory, be identical. To the extent, however, to which specific interference operates, there should be poorer recall for repeated pairs, because these are the OL pairs for which there is a corresponding A-C pair producing maximal interference.

One further point should be made about the ML technique as used in the RI experiment. By using the ML design, it is possible to control for degree of interpolated learning of A-C and C-D pairs, which is extremely difficult to do in the usual unmixed-list (UL) design. In the UL design, the entire interpolated list consists of either A-C or C-D pairs. The A-C list is usually learned more slowly than the C-D list due to negative transfer. If equal trials are given on the A-C and C-D list, the C-D interpolated list is much better learned, which leads to unequal interpolated learning for A-C and C-D lists. What is usually done to avoid this problem is to have the interpolated lists learned to the same criterion. This means that the two lists will be learned for unequal trials, which could lead to differences in the unlearning of the OL pairs. In addition, the distribution of item strengths will also be affected, because even in the A-C list some pairs will be learned early and

then receive more reinforcements than are possible for the C-D pairs. In the mixed-list design, even though A-C and C-D pairs will be learned at different rates, the nature of the design allows a control over degree of interpolated learning (even from the point of view of specific interference). This can be seen by examining a particular R OL pair and a particular NR pair taken from Table 1.

For example, the  $A_1-B_1$  R pair is interfered with by one A-C pair ( $A_1-B_5$ ) and three pairs which are C-D with respect to it ( $A_2-B_6$ ,  $A_5-B_7$ ,  $A_6-B_8$ ). The  $A_3-B_3$  NR pair is also interfered with by these same three C-D pairs. The only difference between the R and NR pairs is that the interpolated pair,  $A_1-B_5$ , is A-C in respect to the R pair and C-D in respect to the NR pair. Since it is only one and the *same* interpolated pair which is responsible for the possible differential interference between the R and NR pairs, it is clear that degree of interpolated learning is controlled with the ML design.

In the experiments as actually performed, there were 10 OL pairs (five R and five NR) and 10 interpolated pairs. Each R pair is interfered with by one A-C pair and nine C-D pairs, while each NR pair is interfered with by 10 C-D pairs. It might be argued that the difference of one versus zero A-C pairs, and nine versus 10 C-D pairs is not sufficient to produce differential interference between R and NR pairs. It should be realized, however, that exactly the same ratios of A-C and C-D interference obtain when the standard UL is employed, and A-C versus C-D interference is studied.

## B. EXPERIMENT I

### 1. Method

All *Ss* learned the same first list. The experimental *Ss* then learned a second list in which half the items were A-C in respect to the first list, and half were C-D. The control group learned no interpolated list. All learning was by the method of anticipation, with the pairs exposed in a Lipman-type memory drum. The original list was presented for seven trials at a four-second rate: two seconds for the stimulus alone and two seconds for the stimulus and response. The pairs were presented in a different random order each trial, and there was no intertrial interval. Twenty minutes after the end of the seventh trial, all *Ss* were tested for recall of the original list. Two minutes following the end of the first list, the experimental *Ss* learned the second list for seven trials, and filled out the remainder of the 20-minute interval by working on an arithmetical-reasoning task. The control group *Ss* rested for two minutes after original learning and spent the remainder of the interval

working on the arithmetic task. For all groups, presentation of the pairs at recall was again at the four-second rate.

All lists consisted of 10 paired associates. The stimuli were nonsense syllables ranging from 40 to 60 per cent in association value as given by the Glaze list in Stevens' Handbook (5), and the responses were four- or five-letter, one-syllable nouns with a frequency of occurrence of five to 17, as given in the Thorndike-Lorge (6) general count. All Ss learned the same original list. The interpolated lists were constructed from the 10 stimuli of the OL, 10 new stimuli, and 10 new responses which met the criterion used for the first list. Each pair in the first list had a corresponding C-D pair which appeared in the IL when it was designated as a NR pair, and a corresponding A-C pair which appeared in the IL when it was an R pair. There was thus a pool of 10 A-C pairs and 10 C-D pairs from which the interpolated lists could be constructed. There were eight interpolated lists, and each was constructed by selecting five A-C pairs and five C-D pairs. Each original pair was A-C with respect to four interpolated lists, and C-D with respect to the remaining four. It was hoped that this procedure would effectively counterbalance differences in OL pairs which could contaminate the results.

Subjects were summer school-undergraduates at the University of Pennsylvania. Sixteen Ss were used in the experimental group, two Ss being assigned to each interpolated list. Ten Ss were used in the control group.

## 2. Results

The data for Experiment I are given in Table 2. Learning scores refer to the mean number of items correctly anticipated on the last learning trial.

TABLE 2  
MEAN LEARNING AND LOSS SCORES OF EXPERIMENT I

Group	Mean learning List 1	Mean learning List 2	Mean loss List 1
Repeated Experimental	6.8	3.2 7.0	3.3 3.4
Nonrepeated Control	4.1	3.6 —	3.7 —
			1.7 1.8 —

The loss scores are obtained for each subject by subtracting the number of items he correctly recalled from the OL after interpolation, from the number of items correct on the last learning trial.

The experimental group learned more than did the control group (by a *t* test for unrelated samples,  $p < .01$ ). Since the groups were not equated

for original learning, the control and the experimental groups were compared on loss scores. Table 1 shows a large difference in loss scores between experimental and control groups ( $p < .01$ ). It seems safe to conclude that there was RI for the experimental group. The critical comparison is between the recall of R and NR pairs of the experimental group. Because each *S* learned five R and five NR pairs, it was possible to use each *S* as his own control and to compare the learning and recall of these pairs within the same subject. The difference between the learning of repeated and nonrepeated pairs was not significant ( $t$  for related samples = 1.24), and neither was the difference in loss scores ( $t = .22$ ). As far as the present experiment is concerned, it would seem that the RI produced can be accounted for by general rather than specific interference.

It is also of interest to examine the learning of the A-C and C-D pairs in the interpolated list, because Twedt and Underwood (7) have shown specific transfer effects when using the ML design. In the present experiment, there was a small difference between A-C and C-D learning, but the difference was not significant ( $t$  for related samples = .83). This data is not necessarily contradictory to the data reported by Twedt and Underwood because different stimuli are involved in the A-C and C-D pairs, and therefore, item differences are not controlled. In the Twedt and Underwood experiment, which was concerned with transfer, there was adequate control for items.

Although Experiment I shows no difference in recall of R and NR pairs, it would be premature to take this result as conclusive support for a non-specific interference theory because that would be tantamount to accepting the null hypothesis as proven. If, however, a number of further experiments with varied conditions revealed the same finding of no difference, confidence in the nonspecific interference hypothesis would be strengthened.

## C. EXPERIMENT II

### 1. Method

Experiment II was an attempt to confirm the results of Experiment I by the use of a slightly different procedure. Following the learning of the 10 A-B pairs, an experimental group learned only five A-C pairs. Thus, the experimental group in Experiment II is treated in the same fashion as the ML *Ss* in Experiment I, except that the C-D pairs are eliminated from the IL. A control group was given the OL to learn for the same number of trials as the experimental group and then worked on an arithmetic task for 20 minutes, at which point the controls were tested for OL recall. The experimental group, when finished with IL learning, worked on the arithmetic



task until 20 minutes had elapsed from the end of OL, and were then tested for recall of OL.

Two experiments were performed using this procedure. In the first experiment (IIa) the material was exposed manually on white index cards, and timing was by metronome. The standard four-second rate of exposure was employed with an eight-second intertrial interval. Eight trials were given for both OL and IL learning. The second experiment (IIb) was identical except that the material was exposed by a memory drum. The lists were the same as had been used in Experiment I, and again the anticipation method was employed.

The *Ss* in Experiment IIa were 16 Yeshiva University graduate students. An attempt was made to match *Ss* so that for each *S* in the experimental group there was an *S* in the control group who learned the same number of items on the original list. Subjects in Experiment IIb were 40 undergraduates at Stern College for Women, 24 in the experimental group and 16 in the control group. No attempt was made to match *Ss* in this experiment.

## 2. Results

The results for Experiment IIa are given in Table 3. The mean loss score is significantly greater for the experimental group ( $p < .005$ ), showing that the five interpolated pairs produced RI. The mean learning scores for the repeated and the nonrepeated A-B pairs are not significantly different ( $t$  for related samples = .25), and neither are the mean loss scores ( $t = .40$ ). Experiment IIa thus confirms the original finding of no difference in retention between repeated and nonrepeated pairs.

TABLE 3  
MEAN LEARNING AND LOSS SCORES OF EXPERIMENT IIa

Group	Mean learning	Mean loss
Repeated Experimental	7.4	1.1
Nonrepeated Control	7.5	1.3

The results for Experiment IIb are given in Table 4. Although the mean loss score for the experimental group was higher than that for the control group, the difference was not significant ( $t = 1.58$ ,  $df = 38$ ). An item analysis of the data showed that OL pairs which had been correctly anticipated on more than two trials showed no more forgetting than control pairs which had been anticipated the same number of times. Experimental pairs which had

been anticipated one or two times, however, were forgotten more readily than corresponding pairs in the control group. The analysis, therefore, was performed with pairs of OL strength 1 and 2. Table 5 presents the data for these pairs. The subjects of the experimental and the control groups did not differ significantly in the number of pairs of strength 1 or 2 learned ( $t = .40$ ,  $df = 38$ ). There was, however, a significant difference in the recall of these pairs for the two groups ( $t = 3.00$ ,  $df = 38$ ). It seems clear that RI was produced for pairs of OL strength 1 and 2.

TABLE 4  
MEAN LEARNING AND LOSS SCORES OF EXPERIMENT IIb

Group	Mean learning		Mean loss	
Repeated Experimental	6.3	3.2	1.2	.6
Nonrepeated Control	6.4	3.1	.6	.5

TABLE 5  
MEAN LEARNING AND LOSS SCORES OF EXPERIMENT IIb PAIRS OF OL  
STRENGTH ONE AND TWO

Group	Mean learning		Mean loss	
Repeated Experimental	2.9	1.5	2.1	1.0
Nonrepeated Control	2.4	1.4	1.1	1.0

The repeated and the nonrepeated pairs of OL strength 1 and 2 were then examined for learning and retention. The analysis of these pairs showed no difference in learning ( $t = .02$ ,  $df = 23$ ) and no difference in recall ( $t = .03$ ,  $df = 23$ ). Experiment IIb, then, confirms the previous results by showing that both repeated and nonrepeated pairs are equally affected by the interpolated list.

#### D. EXPERIMENT III

##### 1. Method

Experiment III was designed as a replication study of Experiment I. Because data from previous experiments were available, it was possible to design the experiment in a way more sensitive to possible differences in the recall of repeated and nonrepeated pairs than were Experiments I and II. Two major differences were incorporated. First, subjects were matched in the different conditions on the basis of total anticipations throughout learning,

instead of on the number of items correct on the last learning trial. From a number of previous experiments, it was ascertained that the correlation between total anticipations and recall was higher than the correlation between number of items correct on the last trial and recall; therefore, total anticipations were used to match subjects.

The second major change involved the ML. In Experiments I and II, eight different IL lists were used, each list having a different combination of repeated and nonrepeated IL pairs. In the present experiment, only two different interpolated lists were used. Each list had five A-C pairs and five C-D pairs. The pairs that were A-C for one list were C-D for the other, and *vice-versa*. By use of this technique, it was possible to match subjects so that each subject who had interpolated list A was paired (on the basis of total A-B anticipations) with a subject who had interpolated list B. This procedure was thought to be more effective in equating each item under repeated and nonrepeated conditions.

The instructions to the subjects were identical to those used in Experiment I. The lists were exposed on a Stoelting memory drum at a 2:2 rate with a six-second intertrial interval. Nine OL trials were given, and on the last trial, only the stimulus members of each pair were presented. There was a three-minute interlist interval. The interpolated list was presented for the same number of trials and in the same way as for the OL. The experimental group worked on an arithmetic task from the end of IL learning to recall; the control group worked on this task after OL. There was a 20-minute interval between the end of OL and recall.

Subjects were 24 Bard College students, 16 in the ML group and eight in the control group. Subjects were matched by the method discussed above.

## 2. Results

Data for Experiment III are presented in Table 6. There was no difference in the learning scores for experimental and control groups, but a significant difference was obtained when the loss scores were compared ( $p < .01$ ). RI for the ML group having been established, the repeated and the nonrepeated pairs were compared. By use of a  $t$  test for related samples, the repeated and nonrepeated pairs did not differ in original learning ( $t = 1.69$ ) or in loss scores ( $t = 1.47$ ). Thus, none of the experiments reported shows conclusive evidence for a specific interference effect.

Even though the difference between repeated and nonrepeated pairs was not significant, the data show a decided trend toward superiority in recall of the NR pairs. In terms of per cent loss, the repeated pairs show a decrement

TABLE 6  
MEAN LEARNING AND LOSS SCORES OF EXPERIMENT III

Group	Mean learning		Mean loss	
Repeated		3.6		1.2
Experimental	7.5		1.9	
Nonrepeated		3.9		.7
Control	7.6		.1	

in recall of 33 per cent, while the nonrepeated pairs show a loss of only 18 per cent. This difference seems too large to be ignored. In an attempt to obtain more information from the present data, an item analysis was performed, and the results are presented in Figure 1. All pairs correctly anticipated a given number of times were pooled, and the percentage of these pairs recalled was recorded. The graph describing percentage recall as a function of number of OL anticipations was then plotted for the repeated and the nonrepeated pairs. By way of comparison, the same analysis was also performed for the data of Experiment I.

The results indicate that the effect of specific interference is related to the strength of learning of OL pairs. The curves for both Experiment I and Experiment III show that as OL increases, the advantage in recall for nonrepeated items increases. Included in Figure 1 is the curve for the control group of Experiment III (the control from Experiment I was not included because the subjects in that group did not learn as well as the experimental subjects). Comparison of the nonrepeated and the repeated curves with the control-group curve shows that the recall of both repeated and nonrepeated pairs is affected by the interpolated list when OL learning is poor; with higher degrees of learning the nonrepeated pairs become impervious to interference, while the repeated pairs still show the effects of interpolation.

#### E. EXPERIMENT IV

The preceding analysis suggested that specific interference occurs when pairs are well learned. A problem remains with the interpretation of this result. In a given experiment, the pairs which receive many anticipations will tend to occur for *Ss* who are the better learners. Therefore, it is difficult to know whether the specific interference which occurs for items of high OL strength reflects something about item strength *per se*, or whether it reflects differences in the way poor and good learners are affected by interpolated material. It could be the case, for instance, that good learners attempt to utilize the A-B associations in learning the second list, and that this leads to specific interference only for these subjects. The interpretation of general and



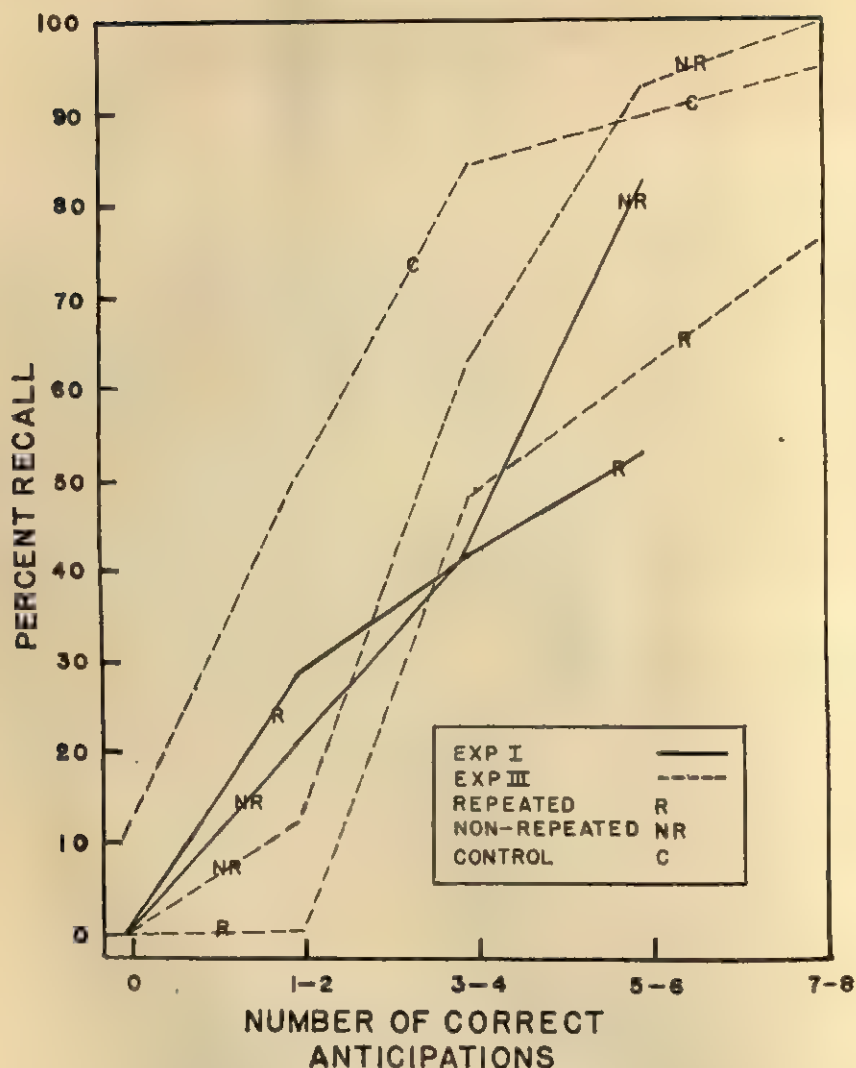


FIGURE 1

PER CENT RECALL AS A FUNCTION OF CORRECT OL ANTICIPATIONS

specific interference would certainly depend on whether this interference is related to item strength or to individual differences. This point can be settled fairly easily by manipulating degree of OL learning as an independent variable.

If different groups of *Ss* are given different numbers of OL trials, then

subject differences and item-strength differences can be teased apart, whereas they are confounded in the previous analysis. If specific interference increases with number of OL trials, then item strength would seem to be crucial as the determinant of specific interference.

### 1. Method

Three groups were employed. Groups A, B, and C practiced OL for two, four, and eight trials respectively. The IL was practiced for four trials for the three groups. The lists used were identical to those used for the ML group in Experiment III, and the material was presented manually on white index cards. The stimulus was shown alone for three seconds; then the card was turned over and the stimulus and response members were shown for three more seconds. The intertrial interval was 10 seconds during which time E shuffled the cards. A one-minute interval separated OL and IL. At the end of IL, subjects worked on an arithmetic problem for five minutes and were then tested for OL recall.

Thirty Brooklyn College students were used as subjects, 10 in each group. No attempt was made to match groups.

### 2. Results

As can be seen from Table 7, the mean loss scores in Group A are practically identical for repeated and nonrepeated pairs, while in groups B and C a difference appears; however, only in the case of Group B is the difference significant ( $t = 2.2$ ,  $df = 9$ ). These results support the hypothesis that specific interference occurs when the OL items are well learned. Of course, when the strength of OL terms becomes very great, then little RI would be expected, and this should diminish the difference between the recall of repeated and nonrepeated pairs. This factor is probably operating in Group C of the present experiment and can also be seen in Figure 1, where the difference in recall between repeated and nonrepeated pairs diminishes when these pairs are of strength 7-8.

### F. DISCUSSION

The results of Experiment IV establish that under some conditions direct evidence for specific interference can be obtained. The fact that, in the previous experiments, it was not possible to demonstrate specific interference, and the fact that in any given experiment specific interference is most evident for strong items seem to indicate that both specific and nonspecific effects operate in RI. In general, the results indicate that weak items are lost due to a nonspecific effect, and strong items due to a specific effect.

TABLE 7  
MEAN LEARNING AND LOSS SCORES OF EXPERIMENT IV

Group	Mean learning	Mean loss
A. Two-trial ML		
Repeated	2.0	1.1
Nonrepeated	2.4	1.2
B. Four-trial ML		
Repeated	3.3	1.7
Nonrepeated	3.2	.6
C. Eight-trial ML		
Repeated	4.4	1.0
Nonrepeated	4.7	.4

A problem which arises with the demonstration of specific interference is that these results seem contradictory to the findings of Runquist reported at the start of this paper. If there is a specific effect of A-C on A-B, then why does the strength of A-C seem irrelevant to the recall of A-B?

A possible resolution to this difficulty would be to assume that the specific interference effect is produced by unlearning. That is, one could assume that the presence of a specific A-C pair is necessary for the unlearning of A-B. The degree to which A-C is learned need not be critical in the loss of A-B because unlearning may be related to the attempt to learn A-C and not to the degree to which A-C is actually learned. The results reported in the present paper could be taken to mean that the nonspecific effect exerts its influence chiefly on the weak items and obscures the specific effect for these items. As degree of OL learning increases, any RI produced will tend to be primarily caused by unlearning and will result in greater specific interference.

The nonspecific effect could be attributed chiefly to recall interference (that is, to some sort of competition at the point of recall). The implication of this would be that, in PI, interference should be completely nonspecific, because PI is thought to be a pure case of recall interference. Runquist has shown that the strength of the specific competing response is not a factor in recall for PI as well as RI; so the assumption that PI is characterized by nonspecific interference is compatible with his finding.

#### G. SUMMARY

Some evidence exists in the literature which suggests that nonspecific interference characterizes both RI and PI. In this study an attempt was made to demonstrate specific interference in an RI design. The basic design involved a mixed interpolated list, where half the items were A-C with respect to the first list and half were C-D. The results of several experiments show specific

interference when OL items are well learned and nonspecific interference when OL items are poorly learned. These results taken in conjunction with other results reported in the literature suggest that unlearning produces specific interference, while recall interference (i.e., interference at the point of recall) produces nonspecific interference. This hypothesis leads to the deduction that PI is characterized by nonspecific interference.

## REFERENCES

1. BILODEAU, I. M., & SCHLOSBERG, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *J. Exper. Psychol.*, 1951, **41**, 199-204.
2. GREENSPOON, J., & RANYARD, R. Stimulus conditions and retroactive inhibition. *J. Exper. Psychol.*, 1957, **53**, 55-59.
3. POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal Learning and Verbal Behavior*. New York: McGraw-Hill, 1961.
4. RUNQUIST, W. N. Retention of verbal associates as a function of strength. *J. Exper. Psychol.*, 1957, **54**, 369-375.
5. STEVENS, S. S., Ed. *Handbook of Experimental Psychology*. New York: Wiley, 1950.
6. THORNDIKE, E. L., & LORGE, I. *The Teachers Word Book of 30,000 Words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.
7. TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. Exper. Psychol.*, 1959, **58**, 111-116.

*Department of Psychology*  
*Graduate School of Education*  
*Yeshiva University*  
*110 West 57 Street*  
*New York 19, New York*



# COMMUNICATION AND LOGIC: VI. EVIDENCE FOR SHOWING THE INFLUENCE OF NEGATIVE THOUGHTS ON COMMUNICATION\*<sup>1</sup>

*Campbell-Ewald Company, Detroit, Michigan*

DANIEL K. STEWART

## A. STATEMENT OF THE PROBLEM

In this paper, the author discusses one aspect of a research project that was concerned with discerning the truth-value of the hypothesis that the communication of ideas is a function of the isometry of validity patterns that exist in the minds of those involved in the communicative process: namely, a speaker and an interpreter (4). However, prior to the investigation of the hypothesis it was necessary to obtain data on the influence of negative ideas on reasoning ability. It is the function of the present paper to present such information.

## B. MATERIALS AND METHODOLOGY

Eight test instruments, "logical questionnaires," were employed. In each of the questionnaires, three validity patterns were used. These patterns were as follows: Modus Ponens (MP), Modus Tollens (MT), and Syllogism (Syll). [See (2, 3).] The subject matter of each test item was neutrally toned and of a general nature (1). Test items were constructed in pairs. Each test item involved a single judgment on the part of the interpreter. The scoring formula, Rights minus Wrongs, was employed. All reliability values were obtained from the formula:

$$r = \frac{n}{n-1} \left[ \frac{\sigma_s^2 - \left( \frac{\bar{x}}{n} \right) (n - \bar{x})}{\sigma_s^2} \right]$$

The test instruments were administered to a total of 822 students enrolled in a Freshmen course in Natural Science at Michigan State University in the Spring Quarter of 1961.<sup>4</sup>

\* Received in the Editorial Office on February 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The data in this paper were secured in a research project sponsored by the National Institute of Mental Health. The project period extended from May 1, 1960, through April 30, 1962.

Four of the test instruments were designed to obtain evidence for the influence of negative ideas, as explicated by propositional forms, and four as explicated by predicative forms.

Each subgroup (i.e., the propositional, and the predicative) contained four tests. Each of the four tests in the propositional subgroup was distinguished from each other by the quantity of items. For example, one test in the propositional subgroup consisted of 30 items. These 30 items were separated into three equal groups of Modus Ponens, Modus Tollens, and Syllogism items. Ten items consisted of the Modus Ponens form (five of these were "positive" items and were paired against five "negative" items); 10 items consisted of the Modus Tollens form (five "positive" items were paired against five "negative" items); and 10 items consisted of the Syllogism form (five "positive" items were paired against five "negative" items).

A second test in this propositional subgroup consisted of 42 items. As with the above test, this test was divided into three equal groups according to the forms Modus Ponens, Modus Tollens, and Syllogism. In this test, however, there were seven "positive" Modus Ponens items paired against seven "negative" Modus Ponens items, seven "positive" Modus Tollens items paired against seven "negative" Modus Tollens items, and seven "positive" Syllogism items paired against seven "negative" Syllogism items.

A third and a fourth test in this propositional subgroup were similarly constructed. Each test contained a total of 54 and 66 items, respectively. The third test paired nine "positive" items per form against nine "negative" items per form, while the fourth test paired 11 "positive" items per form against 11 "negative" items per form.

Exactly the same distinctions were made within the four test instruments within the predicative subgroup. The only difference was the type of form employed. Here, all the items were expressed in predicative form.

The reason for having this range in the number of items in both subgroups was to enable us to select the test that yielded the highest reliability value but which had the smallest number of items per validity pattern.

Two sample pairs of validity patterns are as follows:

### 1. *First Sample Pair*

#### *English*

If Bill sits beneath the lamp, his bald head will shine in Sam's eyes.  
But, if Alice sits in the brown chair, Bill will sit beneath the lamp.

According to this, then, if Alice didn't sit in the trown chair, Bill will sit beneath the lamp.

Question: is the above reasonable or not?

(correct answer: no)

*Validity pattern (propositional Syll)*

$$\begin{array}{r} A \supset B \\ B \supset S \\ \hline \therefore \sim A \supset S \end{array}$$

*English*

If the Sanitary Engineer attends the Elk Club banquet, the Mayor will not win the election. But, if the sidewalks along Main Street are not dirty, then the Sanitary Engineer will attend the Elk Club banquet. This means, therefore, if the sidewalks along Main Street are not dirty, the Mayor will not win the election.

Question: is the above reasonable or not?

(correct answer: yes)

*Validity pattern (propositional Syll)*

$$\begin{array}{r} \sim D \supset E \\ E \supset \sim W \\ \hline \therefore \sim D \supset \sim W \end{array}$$

## 2. Second Sample Pair

*English*

Rare things are expensive. Since things devoted to truth are rare, it follows that things devoted to truth are expensive.

Question: is the above reasonable or not?

(correct answer: yes)

*Validity pattern (predicative Syll)*

$$\begin{array}{r} (x) \cdot T_x \supset R_x \\ (x) \cdot R_x \supset E_x \\ \hline \therefore (x) \cdot T_x \supset E_x \end{array}$$

*English*

Anyone who is alert is not absentminded. Anyone not a professor is an alert individual. According to this, anyone who is not a professor is not absentminded.

Question: is the above reasonable or not?

(correct answer: yes)

*Validity pattern (predicative Syll)*

$$\begin{array}{r} (x) \cdot \sim P_x \supset A_x \\ (x) \cdot A_x \supset \sim F_x \\ \hline \therefore (x) \cdot \sim P_x \supset \sim F_x \end{array}$$

## C. RESULTS

The results are shown in Tables 1-4.

TABLE 1  
STATISTICS FOR TEST THAT COMPARED NINE "POSITIVE" PROPOSITIONAL ITEMS PER FORM  
AGAINST NINE "NEGATIVE" PROPOSITIONAL ITEMS PER FORM ( $N = 132$ )

Validity pattern	Mean	Standard deviation	Reliability
Propositional (positive) forms			
Modus Ponens	7.568	1.818	.716
Modus Tollens	6.576	2.333	.758
Syllogism	6.939	2.421	.820
Propositional (negative) forms			
Modus Ponens	6.750	2.261	.754
Modus Tollens	4.697	2.483	.716
Syllogism	5.311	2.632	.772
Average			.756

TABLE 2  
SE VALUES (AT ONE PER CENT) BETWEEN THE MEANS OF CORRESPONDING PROPOSITIONAL  
VALIDITY PATTERNS COMPARING NINE "POSITIVE" ITEMS PER FORM AGAINST  
NINE "NEGATIVE" ITEMS PER FORM ( $N = 132$ )

Modus Ponens (positive) and Modus Ponens (negative)	3.735	significant
Modus Tollens (positive) and Modus Tollens (negative)	7.311	significant
Syllogism (positive) and Syllogism (negative)	5.985	significant

TABLE 3  
STATISTICS FOR TEST THAT COMPARED 11 "POSITIVE" PREDICATIVE ITEMS PER FORM  
AGAINST 11 "NEGATIVE" PREDICATIVE ITEMS PER FORM ( $N = 98$ )

Validity pattern	Mean	Standard deviation	Reliability
Predicative (positive) forms			
Modus Ponens	10.122	1.342	.607
Modus Tollens	6.755	3.341	.844
Syllogism	8.857	2.199	.707
Predicative (negative) forms			
Modus Ponens	8.694	1.987	.592
Modus Tollens	6.092	3.156	.800
Syllogism	5.745	3.157	.798
Average			.725

TABLE 4  
SE VALUES (AT ONE PER CENT) BETWEEN THE MEANS OF CORRESPONDING PREDICATIVE  
VALIDITY PATTERNS COMPARING 11 "POSITIVE" ITEMS PER FORM AGAINST  
11 "NEGATIVE" ITEMS PER FORM ( $N = 98$ )

Modus Ponens (positive) and Modus Ponens (negative)	5.925	significant
Modus Tollens (positive) and Modus Tollens (negative)	1.426	not significant
Syllogism (positive) and Syllogism (negative)	8.000	significant



## D. CONCLUSIONS

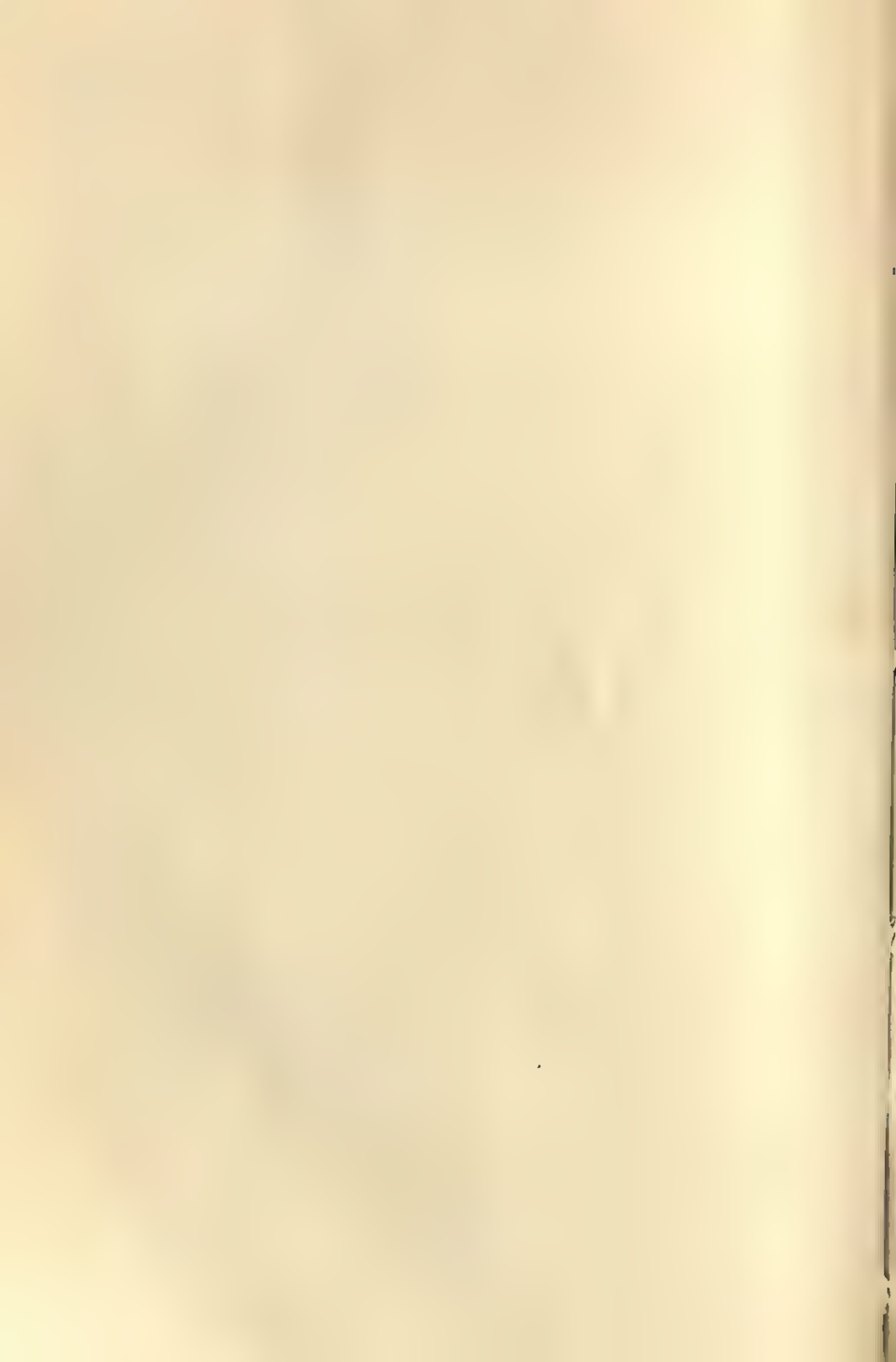
The average reliability values for the four tests in the propositional subgroup—according to ascending quantity of items—are .714, .692, .756 (Table 1), and .788 respectively. The test that paired nine "positive" items per form against nine "negative" items per form was selected for further analysis (Table 2). The results for *propositional* forms indicate that the employment of negative ideas within a given validity pattern has an inhibitory influence on ability to use that pattern.

The average reliability values for the four tests in the predicative subgroup—according to ascending quantity of items—are .702, .730, .801, and .725 (Table 3), respectively. The test that paired 11 "positive" items per form against 11 "negative" items per form was selected for further analysis (Table 4). The reason that the test results for the nine-item comparisons were not used for additional analysis was because the population involved was small (44 students) and highly homogeneous—favoring the more intelligent students. The results for *predicative* forms indicate that negative ideas have an inhibitory influence for the Modus Ponens and Syllogism patterns, but not for the Modus Tollens pattern.

## REFERENCES

1. LEFFORD, A. The influence of emotional subject matter on logical reasoning. *J. Gen. Psychol.*, 1946, **34**, 127-151.
2. STEWART, D. K. The laws of thought?: An hypothesis. *Psychol. Rep.*, 1960, **7**, 229-232.
3. ———. Modus ponens and the experimental basis for tests of logical thinking. *Educ. & Psychol. Meas.*, 1961, **21**, 951-954.
4. ———. Communication and logic: III. Evidence for showing that communication is a function of validity patterns. *J. Gen. Psychol.*, in press.

*Marketing and Research Department*  
*Campbell-Ewald Company*  
*General Motors Building*  
*Detroit 2, Michigan*



## THE EFFECT OF INSTRUCTION ON THE APPEARANCE OF THE AUTOKINETIC EFFECT\*

*Department of Psychology, Boston College*

JOSEPH CAUTELA AND FRANCIS VITRO

### A. INTRODUCTION

Since von Humbolt (9) first noticed the apparent movement of stars and Aubert (1) invented the term Autokinetic Effect (AE), the phenomenon of Autokinesis has been related to many variables. Such variables as size (2), eye movements (4), illumination of the visual field (7), and suggestion (6, 10) have been explored to determine their influence on the AE. Though there have been many studies investigating the influence of suggestion on the duration, direction, and degree of movement of the AE, no experimental studies have been directly concerned with the influence of variation in instructions on whether or not the AE is perceived.

In the usual experiment on AE, the instructions to the *S* either tell the *S* that he will perceive a moving light or that the light might move.

Since the classic study by Sherif (8), it has been assumed that the AE is very readily perceived by all subjects because of properties inherent in the AE situation. According to Sherif (8, p. 220),

The experimental production of the Autokinetic effect is very easy and works without any exception, provided, of course, that the person does not use special devices to destroy the effect. For in a completely dark room a single point of light cannot be localized definitely, because there is nothing in reference to which you can locate it. The effect takes place even when the person knows perfectly well that the light is not moving. These are facts which are not subject to controversy; anyone can easily test them for himself.

These assertions have for the most part remained unchallenged by other investigators. Our own preliminary studies on the AE indicated that the ability to perceive the effect seems to be related to the kinds of instructions given.

While some investigators (3, 5, 7), during their experimentation, have noticed that sometimes the AE is not perceived even when movement is suggested (5, 7), they did not explore this finding further. This study is

\* Received in the Editorial Office on February 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

an attempt to determine how variations in instructions influence ability to perceive Autokinetic Effect.

## B. METHOD

### 1. *Subjects*

The Ss were 100 male undergraduates at Boston College.

### 2. *Apparatus*

The light source was a 15-watt bulb placed in a light-proof box (15"  $\times$  5"  $\times$  9- $\frac{1}{2}$ "). A hole in the box,  $\frac{1}{8}$ " in diameter, allowed the light to escape.

### 3. *Procedure*

There were four different instruction groups of 25 Ss each. All Ss were blindfolded outside the experimental room and given the following instructions:

This is an experiment in perception. I am going to blindfold you and lead you into a dark room. I will direct you to a chair and after two minutes I will ask you to remove your blindfold; then I will give you further instructions.

The S was then led into the experimental (light-proof) room and seated. The autokinetic light source was 83- $\frac{1}{2}$ " away and directly in front of S at about eye level. During a two-minute adaptation period all Ss were told, "After you have removed your blindfold, focus your complete attention on the light which will be directly in front of you." After this initial statement each S received one of the following sets of instructions:

a. *Group I.* "I want you to report to me anything you see as soon as you see it."

b. *Group II.* "I want you to tell me what the light is doing as soon as you notice anything."

c. *Group III.* "I want you to notice the following characteristics of the light: its color, its brightness, whether it goes off and on. As soon as you notice any of the mentioned characteristics tell me what you notice." After the exposure period, the Ss in this group were asked, "Did you see the light move?"

d. *Group IV.* "I am going to present a moving light. I want you to tell me when it moves, and in what direction it moves."

All Ss were exposed to the autokinetic light for three minutes.



## C. RESULTS AND DISCUSSION

A chi-square analysis shows the effect of instruction on the perception or nonperception of the AE to be significant beyond the .001 level. It is interesting to note (see Table 1) that the "Report anything you see . . ." group

TABLE 1  
THE NUMBER OF *Ss* WHO REPORTED MOVEMENT IN EACH GROUP ( $N = 25$ )

Group	Instructions	Yes	No
I	Report anything you see	4	21
II	Tell me what the light is doing	11	14
III	Notice the characteristics	3	22
IV	I am going to present a moving light	22	3

(Group I) and the "Notice the following characteristics . . ." group (Group III) have about the same number of *Ss* who perceived movement. In Group I, the instructions did not suggest that the light was doing anything and, in Group III, the instructions led the *S* to look for nonmovement characteristics.

In groups II and IV, the *Ss* were told that the light was going to do something. In Group II, though movement was not directly suggested, the instructions that the light was doing something apparently increased the likelihood of movement being one of the things the light would be apt to do.

In the study of Graybiel and Clark (3), in which some *Ss* were asked to report any change in the light, fewer than half the *Ss* did not see movement. These instructions most nearly correspond to our "Report what the light is doing . . ." group. Also in our group, a little less than half of the *Ss* did not perceive movement.

The *Ss* in Groups I and II who did not report movement noticed such things as a halo around the light, lines coming from the light, brightness changes, color changes, and size changes.

In Group II, for which the light was supposed to be doing something, more *Ss* than in Group I reported size and brightness changes. In Group III, as expected, there were more size, color, and brightness changes than in any other group. In Group III, no *S* who did not perceive spontaneous movement reported any movement when he was asked the direct question concerning whether or not he saw movement. In Group IV, movement was perceived within a few seconds by almost all *Ss* who reported movement. As has been reported elsewhere (4, 7) most *Ss* who perceive the AE do so within a few seconds after exposure to the autokinetic light. The three *Ss* in Group IV

who did not perceive movement denied that the light was moving in any way.

The results concerning the effect of suggestion on the possible occurrence of the AE seem conclusive. The more the *S* is led to believe that the light is going to do something (or move) the more apt he is to perceive the effect. The AE is not readily perceived by most *Ss* when movement is not suggested. Even when movement is suggested, some *Ss* do not perceive the AE. The above evidence seems to contradict Sherif's belief that the inherent properties of the Autokinetic-Effect situation (no adequate visual frame) somehow force *S* to perceive movement.

#### REFERENCES

1. AUBERT, H. Die bewegungsempfindung. *Arch. f. ges. Psychol.*, 1887, **40**, 457-480.
2. EDWARDS, W. Autokinetic effect as a function of size and brightness. *J. Exper. Psychol.*, 1954, **48**, 391-398.
3. GRAYBIEL, A., & CLARK, B. The autokinetic illusion and its significance in night flying. Report No. 3, Project No. X-148 (*AV-V4-3*), Naval School of Aviation Medicine, U.S. Naval Air Training Bases, Pensacola, Florida, 1945.
4. GUILFORD, J. P., & DALLENBACH, K. M. A study of the autokinetic sensation. *Amer. J. Psychol.*, 1928, **40**, 83-91.
5. HAGGARD, E. A., & ROSE, G. J. Some effects of mental set and active participation in the conditioning of the autokinetic phenomenon. *J. Exper. Psychol.*, 1944, **34**, 45-59.
6. LINTON, H. B. Autokinetic judgements as a measure of influence. *J. Abn. & Soc. Psychol.*, 1954, **49**, 464-466.
7. LUCHINS, A. S. Autokinetic effect and gradations of illumination of the visual field. *J. Gen. Psychol.*, 1954, **50**, 29-37.
8. SHERIF, M. Group influences upon the formation of norms and attitudes. In E. E. Macoby, T. M. Newcomb, & L. Hartley (Eds.), *Readings in Social Psychology*. New York: Holt, 1958. Pp. 219-232.
9. VON HUMBOLT, A. Voyagex aux regions equinoctiaux. *Kosmos*, 1850, **3**, 73-74.
10. WALTER, N. A study of the effects of conflicting suggestions upon judgements in the autokinetic situation. *Sociometry*, 1955, **18**, 138-146.

Department of Psychology  
 Boston College  
 Chestnut Hill 67, Massachusetts

## TWO-COLOR MIXTURES: I. BROAD-BAND FILTERS\*<sup>1</sup>

*Sylvania Electronic Systems, Waltham, Massachusetts*

---

DONALD W. BROWN

---

### A. INTRODUCTION

The first recorded two-color mixture appears to be that of Louis Ducos du Hauron in 1895 [Wall (14)]. Ducos du Hauron describes the production of "a colored sensation as complete as the trichromatic images" which was obtained using only red and blue pigments (the yellow pigment which he had been using in his three-color separation work was eliminated). This phenomenon was restricted in the sense that it could be observed only under weak daylight or by yellow light, such as that obtained from candles, whereas for three-color mixtures the optimum viewing condition was bright daylight.

In 1959 Land published descriptions of extensive investigations of two-color mixtures (8, 9, 10) that employed the following procedure (8):

... two black and white photographic positive transparencies are made in a split-beam camera. . . . One is taken through a Wratten No. 24 filter, passing wavelengths longer than about 585 m $\mu$  and one through a Wratten No. 58 filter, passing wavelengths shorter than about 585 m $\mu$ . These two photographic records will be referred to as the "long record" and the "short record," respectively. The picture is exposed so that the gray scale appears to have the same densities on both records.

The two records are then projected . . . and superimposed exactly. . . . In front of each lens is a pair of polarizers mounted so that the intensity of the light in each beam can be varied separately. In projecting, the long record is projected ordinarily through a red filter (Wratten No. 24) and the short record through a neutral density filter of about 0.3 (when the polarizers are not being used). The half of the projector containing the long record will be referred to hereafter as the "long projector," and the other as the "short projector."

Employing the procedure described above, a full range of color was obtained including yellow (blond hair), blue, red, blue-green, and flesh tones. It was also demonstrated (a) that there is no appreciable time lapse in the appearance of "full-color," (b) that these colors can be observed in white ambient

---

\* Received in the Editorial Office on February 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This work was sponsored by the United States Air Force under Contract No. AF30(602)-2635 monitored by the Human Engineering Laboratory, Rome Air Development Center, Rome, New York.

illumination, and (c) that the range of color perceived is independent of the visual angle subtended by the stimulus. It should be noted that all of these phenomena were demonstrated in a qualitative manner and it seems extremely unlikely that, for example, the independence of color on visual angle would hold at extreme values (8).

From the demonstrations described (plus others which seem less relevant to the present investigation), Land (8, 9, 10) concluded that color perception could not be described as a function of the wavelengths of the energies impinging upon the visual receptor mechanism and that a new coordinate system was required to describe color vision with "natural images"—as opposed to a stimulus spot in a uniform surround. This conclusion (that classical color theory is inadequate to explain the phenomena described) has been examined in detail by Woolfson (16), Judd (7), and Walls (15). These authors concluded that, if consideration is given to the phenomena of "color constancy" (the discounting of the color of the illuminant in judging the color of objects) and to simultaneous and successive contrast (the induction of the complement of a color in a neutral area adjacent to a chromatic area), the phenomena reported by Land (8, 9, 10) can be explained largely within the framework of classical colorimetric theory.

Land's demonstrations and papers (8, 9, 10) stimulated a number of investigators to attempt to replicate the phenomenon and, when successful replication was achieved, to examine the effects of other unexplored variables on the range of colors perceived in two-color mixtures.

Berg and Forkner (1) used both "natural scenes" and a display of square stimulus spots as subject material. They found that the perceived colors were those of the filter (in the single-filter projections), its approximate complement, and shades adjacent to these two colors (which, however, were considerably less saturated). The colors observed were predictable from classical two-color mixture laws, if it was assumed that the complement of the filter hue had been induced. When the two records were projected, using two filters, both the number and the saturation of the colors perceived were dependent on the wavelength separation of the projection filters.

Adaptation (the induction of the complement of the projection filter) was used by Hughes (6) to predict the colors perceived when a "natural scene" was photographed through red and green filters. When this scene was projected with green and red filters over the short and the long projectors, those colors predicted by classical theory (red, yellow, and green) were observed. When the long record was projected through a red filter and the short record with "white" light, reds, yellows, and desaturated blue-greens



were seen. Similarly, predictable colors were perceived with various other filters. Predictions were made on the basis of a shift in the neutral illumination point toward the filter color, in accordance with color constancy. Complements of the projection filter-color were then induced in areas darker than this shifted neutral point.

Bivens (2) examined the relationship between the difference threshold for hue and the minimum separation of projection filters to produce "good" two-color mixtures (Hecht's data were used for the difference thresholds and Land's for the two-color mixture). The data showed a close relationship between the two variables. For a good two-color effect, the filter separation must be about 10 times as great as the difference threshold in that region of the spectrum.

Comparisons among two-color projections, three-color projections, and actual stimulus material with respect to quality and fidelity were made by Burckhardt and Strutt (3). They found little difference between stimuli projected with red and green filters and those projected with a red filter and incandescent light ("white"). Both of these conditions, however, were inferior to the projections of a three-color system, with respect to both quality and accuracy of judged color.

Rushton (13) found that Grassman's laws hold for three-color matches of two-color projections. Two-color projections were made through a red filter (long record) and no filter (short record). A three-color match made to such a two-color mixture was equal to the sum of two separate three-color matches made to each of the components of the two-color mixture (i.e., red and "white"). This three-color match was found to hold, independent of any reported changes in the hue of the stimulus (i.e., changes due to contrast effects). The two-color system was found to leave more colors unrepresented than presented.

Interest in the anatomical locus of two-color mixture effects has led to attempts to produce this phenomenon with presentation of the two components to the two eyes separately. Pastore (11, 12) found that, in addition to two-color mixtures being producible with binocular stimulation, simultaneous contrast effects could also be produced with black and white drawings that were capable of being fused. When fusion occurred while a red filter was over one eye, simultaneous contrast effects were reported. Geschwind and Segal (4) found that a "full range of colors" was obtainable when the two records were presented to two different eyes simultaneously. The authors concluded that the "Land effect," in some cases at least, must take place either at some level higher than the retina or that there must be some effect

of stimulation of one retina, through higher centers, on the sensitivity of the other retina.

## B. OBJECTIVES

The present research was designed to fulfill two primary objectives: (*a*) to collect data in a systematic way on two-color mixtures; and (*b*) to collect the data in (*a*) in such a form as to make possible a precise and quantitative evaluation of the conditions under which varying numbers of hues and saturations can be perceived, and to determine the fidelity of these colors (i.e., the extent to which they agree with the subject matter photographed).

The research is divided into two sections. The first deals with data obtained with broad-band filters; the second, with data obtained with narrow-band filters. Only the broad-band data are presented here.

The broad-band data were collected using three slides photographed through three different filters (red, green, and blue) and combined to yield three different pairs of slides. These slides were projected through nine combinations of the three filters used in photography, plus the condition of no filter. When two filters were used, the long record was always projected through the longer of the two filters. Three modes of response were utilized: (*a*) absolute judgments of hue and saturations; (*b*) a matching procedure in which the subjects matched the stimulus seen with the left eye to one of a number of Munsell chips presented simultaneously, under Illuminant C, to the other eye; and (*c*) a procedure like (*b*) except that the Munsell chips were presented individually.

## C. METHOD

### 1. Apparatus

Three black and white slides (35 mm) were prepared using 11 Munsell color chips as subject matter. Each slide was photographed through a different colored filter (red, green, and blue). Table 1 describes the set of chips. The chips were arranged in two rows of four chips and in one row of three. Three slides were prepared, one each through the following three filters: (*a*) Wratten 47B, 452.7  $m\mu$ , blue; (*b*) Wratten 11, 552.5  $m\mu$ , green; (*c*) Wratten 29, 632.7  $m\mu$ , red. Densities of the stimuli and background on each slide are given in Table 2.

Two Kodak Signet 500 projectors were mounted on a 3'  $\times$  2' piece of plywood to insure steady images on the screen (Da-Lite). The projectors were equipped with irises and filter holders. Sylvania 750-watt DDB lamps were used as light sources (Tristimulus values for the projection lamp and

TABLE 1  
DEFINITION OF STIMULUS CHIPS IN TERMS OF MUNSELL NOTATION, C.I.E. COORDINATES,  
AND WAVELENGTH

Munsell	C.I.E. coordinates		Dominant wavelength
	x	y	
5 BG 5/8	.210	.329	492 m $\mu$
5 B 5/6	.222	.272	486
10 PB 5/8	.257	.221	455
5 P 5/8	.289	.229	561 C
5 R 5/8	.441	.323	612
5 YR 5/8	.483	.395	587
5 Y 8/8	.416	.438	575
10 GY 5/8	.310	.465	551
10 G 5/8	.229	.372	499
White (N 9, 5/ )	—	—	—
Black (N 1/ )	—	—	—

TABLE 2  
STIMULI AND BACKGROUND DENSITIES

Chip	Number 1 452.7 m $\mu$	Number 2 552.5 m $\mu$	Number 3 632.7 m $\mu$
5 BG	1.42	1.12	1.71
5 B	1.25	1.12	1.53
10 PB	1.10	1.17	1.20
5 P	1.17	1.18	1.12
5 R	1.64	1.18	1.02
5 YR	2.03	1.16	1.06
5 Y	1.61	0.66	0.77
10 GY	1.96	1.13	1.66
10 G	1.57	1.11	1.72
White	0.82	0.57	0.69
Black	2.15	1.97	2.06
Background	0.79	0.58	0.69

three filters are available from the author). A 36"  $\times$  24"  $\times$  36" viewing box with the white interior illuminated by an "illuminant C" contained the matching chips. The chips presented to the subjects were those used for stimuli, but each chip included all of the saturation steps below the one photographed (e.g., for stimulus 5 BG 5/8, the chips presented in the viewing box were 5 BG 5/8, 5/6, 5/4, 5/2, 5/1). The distance of the subject's left eye to the screen was 120 inches, while the distance from the right eye to the matching chips was 34 inches. These two distances resulted in equal visual angles being subtended by the stimuli and the matching chips.

## 2. Subjects

One female (VC) and three male employees of the Applied Research Laboratory served as subjects. All subjects had normal color vision.

## 3. Procedure

All subjects underwent training prior to the beginning of the experiment. During training they learned to make absolute judgments of hue and saturation of Munsell chips, according to a modified version of the procedure employed by Helson, *et al.* (5).

Each subject (*S*) was run under every combination of slides and projection

TABLE 3  
WAVELENGTHS OF FILTERS THROUGH WHICH SLIDE COMBINATIONS WERE PHOTOGRAPHED

Combination	$\lambda$ max 1	$\lambda$ max 2
1	452.7 m $\mu$	552.5 m $\mu$
2	452.7 m $\mu$	632.7 m $\mu$
3	552.5 m $\mu$	632.7 m $\mu$

TABLE 4  
WAVELENGTHS OF PROJECTION FILTERS THROUGH WHICH SLIDE COMBINATIONS WERE PHOTOGRAPHED\*

Combination	$\lambda$ max short record	$\lambda$ max long record
1	452.7 m $\mu$	552.5 m $\mu$
2	452.7 m $\mu$	632.7 m $\mu$
3	452.7 m $\mu$	None
4	552.7 m $\mu$	632.7 m $\mu$
5	552.7 m $\mu$	None
6	632.7 m $\mu$	None
7	None	452.7 m $\mu$
8	None	552.5 m $\mu$
9	None	632.7 m $\mu$

\* For each subject, each combination was run with each of the three slide pairs listed in Table 3.

filters. These combinations are given in Tables 3 and 4. Three subjects were run on three response procedures: absolute judgments, matching with all matching stimuli present, and matching with the matching stimuli presented one at a time. One subject (JG) contributed data under the absolute judgment condition only. All data were replicated once for each subject. Thus a typical subject twice received all combinations of slides, projection filters, and response procedures. Stimuli were projected with no ambient illumination, and subjects adapted to the background illumination for 10 minutes prior to the start of an experimental session. In the case of the absolute-judgment procedure, both of *S*'s eyes were adapted to the background provided by the projectors and filters; while for the matching procedure, *S*'s left eyes were adapted to the projection light (25 ft. L) while their right eyes adapted to the illumination inside the matching box (Illuminant "C" at 25 ft. L).

Subjects' responses were recorded in terms of Munsell Hue and Chroma.



#### D. RESULTS

The wavelength (in  $m\mu$ ) and ranges of excitation purity ( $P_e$ ) of the stimulus chips and modal responses (with all response procedures pooled) are presented in Tables 5, 6, and 7.

Table 8 indicates the hues that were perceived on at least five per cent of the trials, with each of the filter combinations employed. Presented also are brief comments that list the hues *not* obtained with each filter pair and give an explanation, where possible, of the reported colors in terms of traditional color-mixture theory when consideration is given to the effects of hues induced by the background color. Figures 1 through 4 are plots of the colors obtained (with four sample pairs of filters) as a function of the densities of the stimuli on the long and short records.

#### E. DISCUSSION OF RESULTS

As can be seen in Tables 5, 6, and 7, the fidelity with which a two-component color system—such as the one employed in this study—can be expected to produce the actual hues and saturations photographed is quite low relative to a three-component system. Of the 27 experimental conditions employed, the greatest number of correct hue responses was four in a series of nine. In addition, and probably contributing to the problem of hue fidelity, is the observation that stimuli frequently tend to become desaturated relative to the original stimulus material.

The other (and perhaps more important) practical consideration—the extent of the range of colors obtainable with a two-color system—seems to receive a more promising answer. Table 8 shows that, with certain combinations of projection filters and the density combinations employed, a full range of colors is attainable. Although only two (green and red, and green and incandescent) of the nine projection conditions employed yielded a full range of colors, it is possible that with different density combinations some (or all) of the other conditions would have yielded these colors. On the basis of this finding, two-component colored displays are shown to have utility, at least in special situations in which the additional saturation to be had with a three-component system is unnecessary.

Also presented in Table 8 are possible explanations, within the framework of classical color-mixture theory, of why these hues were perceived as they were. With the additional well-supported assumption concerning induced colors, it is seen that classical color-mixture theory is adequate to explain all of the hues reported. It should be noted, however, that the explanations are

TABLE 5  
 PURITY ( $P_e$ ) RANGES ASSOCIATED WITH MODAL RESPONSES (IN  $m\mu$ )  
 FOR SLIDE COMBINATION 1

	Chip																					
	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$	$\lambda$	$P_{\odot}$		
	551	40	575	61	486	47	612	37	492	36	455	33	561c	31	499	27	587	68	Background			
Projection filter combination	1	551 575	10-31 18-61	575 30-61	455	4-18	551 575	5-20 11-48	575	11-48	455	8-33	455	8-33	575	11-61	575	11-61	455	18-33		
	2	612	9-37	612	37-60	455	4-33	498c	15-30	612	9-18	455	8-33	455	8-33	612	4-37	612	4-37	498c	15-60	
	3	575 551	18-61 5-20	575 30-61	455	4-26	575	11-61	575	11-48	455	8-33	455	4-33	575	11-61	575	18-61	455	8-70		
	4	612 498c	18-28 15-30	612	37-50	551	5-20	561c	8-25	612	4-18	499 492	13-20 18-27	492	3-18	612	9-37	612	9-37	575	18-61	
	5	612	4-18	561c	15-31	551	5-10	612 498c 561c	4-9 4-5 4-24	587	8-18	551	10-20	551	5-20	612	4-9	561c	4-31	white		
	6	492	3-18	492 499	9-27 13-27	612	18-28	486	7-25	612 587	4-9 8-18	612	18-37	612	9-37	492	3-9 499	486 3-6	7-25 3-6	612	9-70	
	7	486	7-37	455	26-33	575	11-48	486	7-37	561c	4-17	575	11-61	575	11-48	486	7-37	486	7-56	455	8-26	
	8	551	5-10	551	5-43	587	8-18	551	5-10	587	8-18	561c	17-24	561c	8-17	499	3-6	551	5-10	572	17-47	
	9	612 498c	4-37 5-30	612	37-60	587	8-18	498c 612	5-22 4-28	612	4-18	499	6-13 492	499	3-13 3-18	612	9-37	498c	4-30	612	9-70	

TABLE 6  
 PURITY ( $P_e$ ) RANGES ASSOCIATED WITH MODAL RESPONSES (IN  $m\mu$ )  
 FOR SLIDE COMBINATION 2

	Chip																			
	$\lambda$ 551	$P_e$ 40	$\lambda$ 575	$P_e$ 61	$\lambda$ 486	$P_e$ 47	$\lambda$ 612	$P_e$ 37	$\lambda$ 492	$P_e$ 36	$\lambda$ 455	$P_e$ 33	$\lambda$ 561c	$P_e$ 31	$\lambda$ 499	$P_e$ 27	$\lambda$ 587	$P_e$ 67	$\lambda$ 67	$P_e$ Background
1	575	30-61	575	30-61	486	25-37	575	30-61	486	7-37	455	8-33	455	8-18	486	7-37	575	30-61	455	18-70
2	612	4-37 <sup>p</sup>	612	37-50	486	7-50	612	18-37	486	7-40	455	18-33	561c	4-17	486	7-37	612	18-46	561c	31-40
3	black		575	30-61	486	13-50	575	18-61	486	7-37	455	8-33	561c	4-17	486	7-37	575	18-61	455	8-80
4	612	4-18	612	37-60	499	6-20	612	28-37	499	3-13	499	3-13	587	8-35	486	7-37	612	18-46	575	11-61
5	black		561c	8-31	551	10-20 499 6-13	561c	8-24	551	5-10	551	5-20	551	5-20	black		561c	8-31	551	20-60
6	black		499	6-27 492 3-36	612	4-37	499	6-13 492 3-27	498c	5-30	612	9-37	612	4-18	612	4-18 498c 4-15	499	3-20	612	9-70
7	486	7-13	486	25-50 455 26-33	575	18-61	486	7-37	575	18-61	575	11-30	561c	8-17	575	11-61	486	13-50	455	4-50
8	black		551	10-43	561c	8-17	551	10-31	561c	8-31	612	4-18 561c 4-17	575	11-18	561c	4-8	499	6-13	572	11-60
9	black		612	37-60	499	6-13	612	9-37	492	3-18	499	3-6	612	4-18	black		612	9-46	612	9-60

DONALD W. BROWN

TABLE 7  
 PURITY ( $P_e$ ) RANGES ASSOCIATED WITH MODAL RESPONSES (IN  $m\mu$ )  
 FOR SLIDE COMBINATION 3

	Chip																			
	$\lambda$ 551	$P_e$ 40	$\lambda$ 575	$P_e$ 61	$\lambda$ 486	$P_e$ 47	$\lambda$ 612	$P_e$ 37	$\lambda$ 492	$P_e$ 36	$\lambda$ 455	$P_e$ 33	$\lambda$ 561c	$P_e$ 31	$\lambda$ 499	$P_e$ 27	$\lambda$ 587	$P_e$ 67	$\lambda$ Background	$P_e$
1	486	7-45	575	11-18	486	7-45	575	18-61	486	7-45	455	4-8	575	11-48	486	7-45	575	11-61	455	8-70
2	486	7-45	498c	15-42	486	7-45	612	25-50	486	7-45	498c	5-22	498c	5-22	486	13-45	612	18-37	498c	15-50
													561c	17-24						
3	486	7-50	575	11-18	486	7-50	575	11-61	486	7-50	575	11-30	575	11-48	486	13-50	575	18-48	455	4-57
4	551	10-31	575	11-61	551	5-31	612	18-37	551	10-31	587	8-18	612	9-28	499	6-13	612	18-37	575	11-61
									499	6-18										
5	551	5-20	575	11-18	551	5-20	561c	8-31	551	5-20	587	8	561c	4-24	551	5-20	561c	8-24	white	
6	612	9-70	612	4-60	612	4-60	492	3-18	612	9-50	612	4-9	499	3-6	612	9-46	499	6-13	612	18-70
	498c	5-30																		
7	572	17-32	561c	4-45	572	11-32	455	8-33	572	32	561c	4-24	455	4-18	572	32-47	455	8-33	white	
			455	4-40											575	30-61				
8	612	4-9	575	11-30	612	4-9	551	10-20	612	4-18	587	8-18	551	5-20	498c	4-15	551	5-20	white	
					587	8-35													575	11-48
9	499	3-9	575	11-30	551	5-20	575	11-30	551	10-20	612	4-9	612	9-28	551	10-20	612	9-28	white	

Projection filter combination



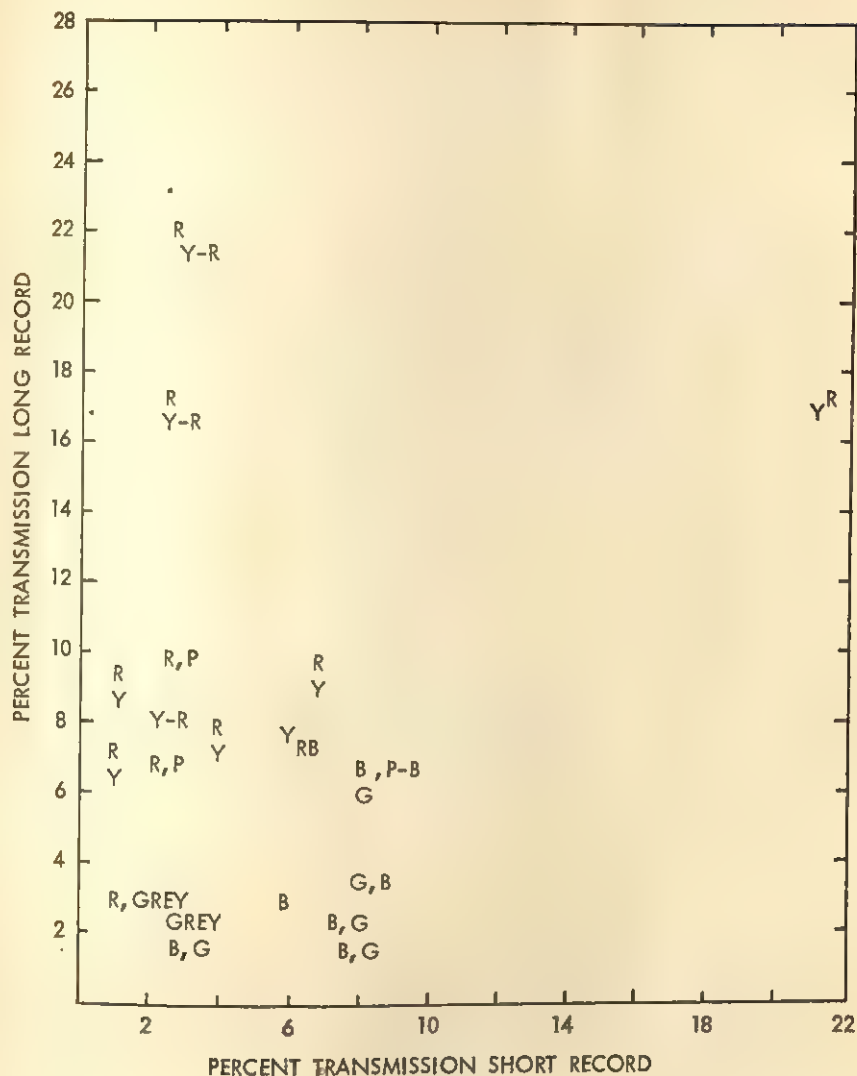


FIGURE 1  
 COLORS REPORTED AS A FUNCTION OF DENSITIES OF LONG AND SHORT RECORDS WITH THE  
 LONG RECORD PROJECTED THROUGH THE RED FILTER AND THE  
 SHORT RECORD WITH NO FILTER

TABLE 8  
HUES PRODUCED BY VARIOUS FILTER COMBINATIONS

Projection filter combination	Perceived back-ground hue (modal responses)	Perceived stimulus hues	Comment
1) blue & green (453 & 533 $m\mu$ )	blue (455 $m\mu$ )	a) green (551 $m\mu$ ) b) blue (455, 486 $m\mu$ ) c) yellow (575 $m\mu$ )	Reds and purples missing. a & b) Traditional color mixture theory predictions. c) A grey spot would take on the hue of the complement of the background. Because the background is 455 $m\mu$ the induced color would be yellow of about 570 $m\mu$ .
2) blue & red (453 & 633 $m\mu$ )	purple (561c $m\mu$ )	a) red (612 $m\mu$ ) b) blue (455, 486 $m\mu$ ) c) purple (498c, 561c $m\mu$ )	Yellows and greens missing. All colors predicted by traditional color mixture theory.
3) blue and incandescent (453 & 583 $m\mu^*$ )	blue (455 $m\mu$ )	a) yellow (572, 575 $m\mu$ ) b) blue (455, 486 $m\mu$ ) c) purple (561c $m\mu$ )	Reds and greens missing. All traditional predictions.
4) green & red (533 & 633 $m\mu$ )	yellow (575 $m\mu$ )	a) red (612 $m\mu$ ) b) green (551, 499 $m\mu$ ) c) yellow (575 $m\mu$ ) d) blue (492, 486 $m\mu$ ) e) purple (561c $m\mu$ )	a, b & c) Red, green and yellow predicted by traditional mixture theory. d) Blue induced by the yellow background in a spot in which red cancels most but not quite all the green; i.e., yellow induces 474 $m\mu$ which mixes with some green to give 486 and 492 $m\mu$ . e) Purple: yellow background induces 474 $m\mu$ which mixes with some projected red to produce purple.
5a) green and incandescent (553 & 583 $m\mu$ )	green (551 $m\mu$ )	a) green (551 $m\mu$ ) b) purple (561c $m\mu$ ) c) blue (499 $m\mu$ )	No red, yellow. a) Green expected on basis of law of intermediates. b) Purple produced by mixture of a little red-yellow (583 $m\mu$ ) with the induced complement (551c $m\mu$ ) of the background (551 $m\mu$ ). c) Blue: induced purple (561c $m\mu$ ) plus projected green (553 $m\mu$ ) produces blue.

TABLE 8 (continued)

Projection filter combination	Perceived back-ground hue (modal responses)	Perceived stimulus hues	Comment*
5b) green and incandescent (553 & 583 $m\mu$ )	yellow (572-575 $m\mu$ )	a) green (551 $m\mu$ ) b) yellow (575 $m\mu$ ) c) purple (561c $m\mu$ ) d) blue (499 $m\mu$ ) e) red (612 $m\mu$ )	a & b) Green and yellow expected on basis of law of intermediates. c) Purple produced by mixture of a little red-yellow (583 $m\mu$ ) with the induced complement (475 $m\mu$ ) of the yellow background. d) Blue produced by mixture of a little yellow (553 $m\mu$ ) with the induced complement (475 $m\mu$ ) of the yellow background. e) Red: yellow-red (583 $m\mu$ ) seen in a yellow background (572-575 $m\mu$ ) would appear more red due to contrast effects.
6) red and incandescent (633 & 583 $m\mu$ )	red (612 $m\mu$ )	a) red (612 $m\mu$ ) b) yellow (587 $m\mu$ ) c) blue (499, 492 $m\mu$ ) d) purple (498c $m\mu$ )	Green missing. a & b) Red and yellow are predicted by the law of intermediates. c) Blue (492 $m\mu$ ) is the approximate complement of red (612 $m\mu$ ); therefore could be induced by the background in a neutral spot. Blue (499 $m\mu$ ) could be produced by a mixture of a little yellow (583 $m\mu$ ) from the lamp with blue of 490 $m\mu$ induced by the background. d) Purple (498c $m\mu$ ) produced by a mixture of induced blue (490 $m\mu$ ) and some projected red in the same spot.

\* The incandescent lamp had a wavelength of 583  $m\mu$  and a purity of 50 per cent.

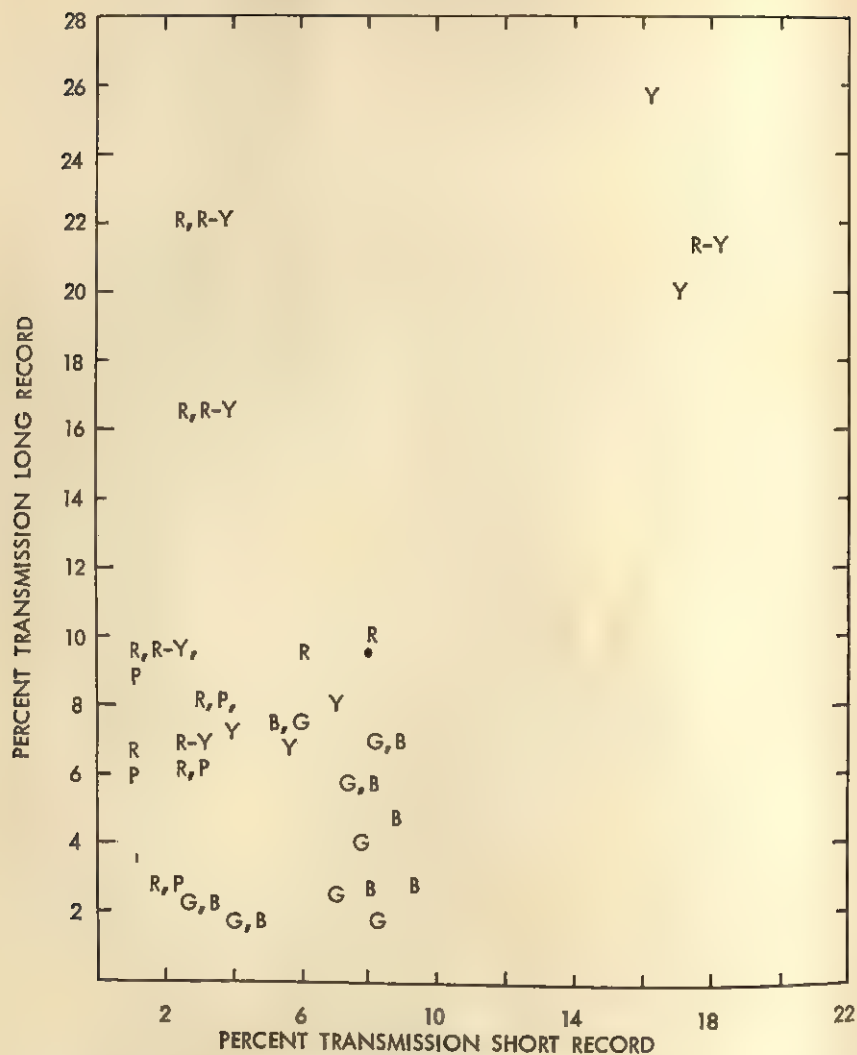


FIGURE 2  
 COLORS REPORTED AS A FUNCTION OF DENSITIES OF LONG AND SHORT RECORDS WITH THE  
 LONG RECORD PROJECTED THROUGH THE RED FILTER AND THE  
 SHORT RECORD THROUGH THE GREEN FILTER



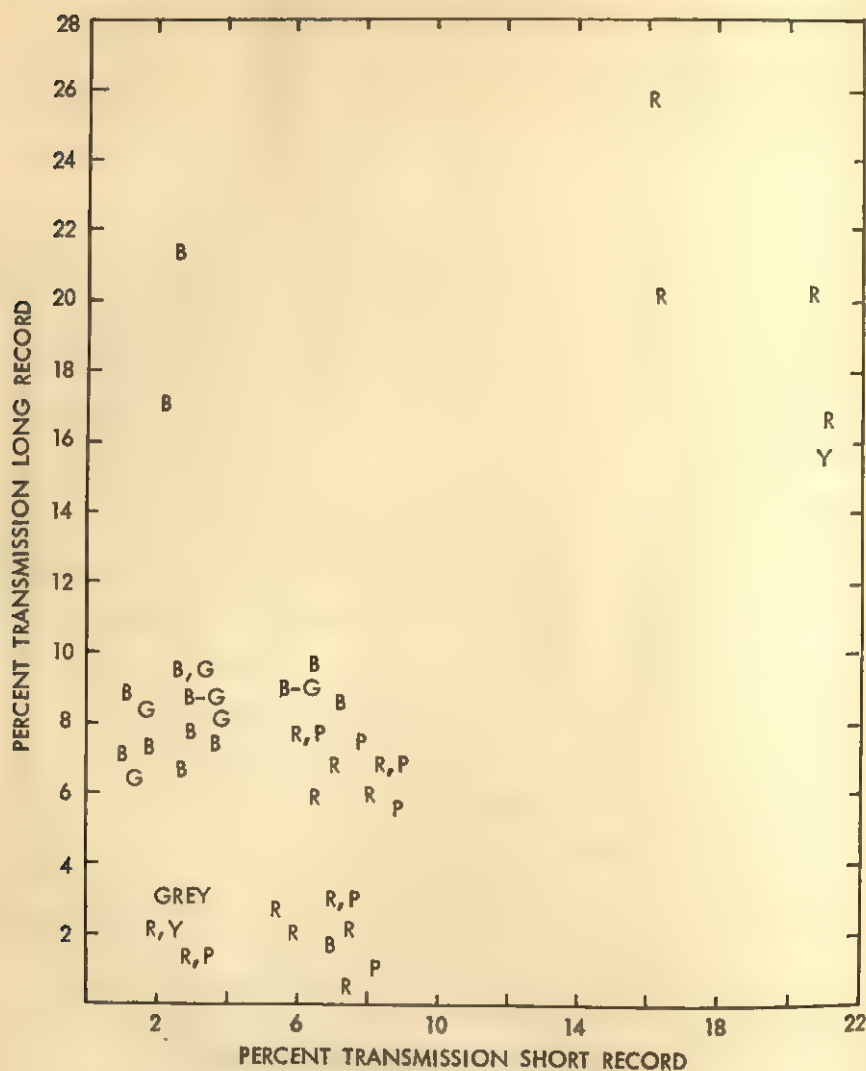


FIGURE 4

COLORS REPORTED AS A FUNCTION OF DENSITIES OF LONG AND SHORT RECORDS WITH THE LONG RECORD PROJECTED THROUGH NO FILTER AND THE SHORT RECORD THROUGH THE RED FILTER

not intended to be complete; that is, there may be, and probably are, other explanations equally valid to explain any given perceived color.

Figures 1 through 4 show that, for the most part, the data from the present study agree with Land's plot of perceived hues as a function of transmission of the long and the short records (8). In a few cases, hues appear on the "wrong" side of the neutral line that represents equal densities; however, the distance from the "correct" position is always small.

#### REFERENCES

1. BERG, S., & FORKNER, P. J. Note on recent demonstrations of color mechanisms. *J. Opt. Soc. America*, 1960, **50**, 394.
2. BIVENS, L. W. Color induction and hue discrimination. *Science*, 1961, **134**, 732.
3. BURCKHARDT, C., & STRUTT, M. J. O. Experiments with E. H. Land's two-color projection. *I.R.E. Transactions on Broadcast & Television Receivers*, 1961, BTR-7, 34-40.
4. GESCHWIND, N., & SEGAL, J. R. Colors of all hues by binocular mixing of two colors. *Science*, 1960, **131**, 608.
5. HELSON, H., JUDD, D. B., & WARREN, M. H. Object color changes from daylight to incandescent filament illumination. *Illum. Eng.*, 1952, **47**, 221-233.
6. HUGHES, W. L. Some color slide and color television experiments using the Land technique. *I.R.E. PGTR*, 1960, BC-6, 29-33.
7. JUDD, D. B. Appraisal of Land's work on two-color primary color projections. *J. Opt. Soc. America*, 1960, **50**, 254-268.
8. LAND, E. H. Color vision and the natural image, Part I. *Proc. Nat. Acad. Sci.*, 1959, **45**, 115-129.
9. ———. Color vision and the natural image, Part II. *Proc. Nat. Acad. Sci.*, 1959, **45**, 636-644.
10. ———. Experiments in color vision. *Sci. Amer.*, 1959, **200**, 84-89.
11. PASTORE, N. Color phenomena. *Sci. Amer.*, 1960, **202**, 168.
12. ———. Color phenomena. *Science*, 1960, **131**, 1400.
13. RUSHTON, W. A. H. The eye, the brain, and Land's two-color projections. *Nature*, 1961, **189**, 440-442.
14. WALL, E. J. *The History of Three-Color Photography*. Boston: American Photographic, 1925.
15. WALLS, G. L. Land! Land! *Psychol. Bull.*, 1960, **57**, 29-48.
16. WOOLFSON, M. M. Some new aspects of color perception. *IBM J. Res. & Devel.*, 1959, **3**, 312-325.

*Sylvania Electronic Systems*  
*Applied Research Laboratory*  
 40 Sylvan Road  
 Waltham 54, Massachusetts

## INTRA-SERIES PRIMACY AND THE ORIENTING REFLEX OF THE GSR\*

*Cornell University Medical College; The New York Hospital, Payne Whitney Psychiatric Clinic; and Hunter College of the City University of New York*

---

R. HAYES, A. BRONZAIT, M. BLOCH, AND L. WELCH

---

### A. INTRODUCTION

Many investigators are aware of the literature on the GSR as an alerting or orienting response to a novel or startle stimulus (1, 2, 3, 4, 7, 8, 9, 14). In 1961, Razran (12) reviewed recent Soviet literature and referred to many experiments devoted to the study of the orienting reflex and said: "There is . . . no doubt that the OR is a prime focusing center in current Soviet psychophysiological and psychological research and thought." Razran further commented: "The total material of the OR . . . does not lend itself readily to summing up and discussing through sample experiments." Under the heading "Reactional primacy and holistic specificity," Razran then discussed the orienting reflex in terms of sets of characteristics. He stated that "the OR is the organism's normal first reaction to any adequate normal stimulus or, in terms of the stimulus, it is the organism's normal reaction to a normal *novel* stimulus." Razran also pointed out that novelty includes change and that it extends broadly to any discriminable changes in stimulus attributes. He noted that a dog manifests a clear-cut OR reaction when shock is changed discriminately in intensity or locus. Russians treat the GSR as one of a variety of physiological components of the OR. Razran further makes the point that:

On the other hand, it should be noted that the OR is specific only to the novelty and change-characteristics of the stimuli that evoke it, but not to their intrinsic nature. In general it might be said that, unlike the alimentary or defensive or sexual reaction patterns of the organism, the OR pattern does not really "manage" the stimuli that come to it, but merely reacts to their presence, its reaction thus being more preparatory than consummatory, and more preadaptive than adaptive.

In our laboratory we always allow a substantial period, usually 20 minutes, for Ss to become accustomed to being "hooked up" with the apparatus, the general appearance of the laboratory, and the Es. We make it a practice to

---

\* Received in the Editorial Office on February 26, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

acquaint *Ss* with the purpose of the experiment and with the procedure that will follow. These general precautions are usually considered sufficient to eliminate most of the characteristics of the stimulus that would be related to "startle" or surprise. Thus, responses differ as a function of stimulus value and individual differences.

In the course of most of our experiments, stimuli are presented in sequence. Randomizing is effected as a safeguard against sequential effects. Despite randomization, if several different types of stimuli are used, ordinality must ensue. This is true for all stimuli and, within the overall sequence, for each type of stimulus; e.g., if three each of five different types of stimuli are used, the presentation resolves into a series from first to 15th and, for each type, from first to third. Reasoning this way, we decided to test the hypothesis that order of presentation was critical with regard to overall "firstness" or "primacy," and "firstness" of each type of stimulus. It was expected that responses to stimuli other than the first stimulus would fall off rapidly.

## B. METHOD

### 1. *Subjects*

*Ss* were 25 female psychology students between the ages of 17 and 50. Participation in the experiment was a course requirement.

### 2. *Apparatus*

The psychogalvanometer used for measuring GSR was a custom-made pathometer capable of measuring changes in skin resistance within about 50 ohms. Responses were recorded on a Texas Instrument Rectiriter which yielded a paper tape recording.

### 3. *Procedure*

The experiment was conducted in a New York Hospital laboratory about a quarter of a mile from the college.

*Ss* were told (*a*) that underway was a study of individual differences in sweat-gland activity; (*b*) that 15 multiple-choice questions (each on a separate card) would be shown to them; (*c*) that some of the questions would be attitude questions for which there were no right or wrong answers, while others pertained to current events or mathematics and, therefore, could yield correct answers (see Table 1); and (*d*) that each question was to be answered with a letter-alternative (*a*, *b*, or *c*; or *a*, *b*, *c*, or *d*) as soon as possible.

Upon the basis of manifest content, questions were designed to represent five groups. These groups were as follows: (*N*) neutral questions, (*CE*)



TABLE 1  
STIMULUS QUESTIONS IN ORDER OF PRESENTATION

Question	Manifest content
1. I think pastels are more pleasant than the strong colors. (a) Yes (b) No (c) Cannot say	Neutral
2. Which one of the following is not a member of Kennedy's cabinet? (a) Goldberg (b) Udall (c) Williams (d) McNamara	Current events
3. Generally speaking, tests make me nervous. (a) Yes (b) No (c) Cannot say	Test-anxiety
4. A great many tests are unfair. (a) Yes (b) No (c) Cannot say	Test-anxiety
5. Premarital intercourse is definitely wrong. (a) Yes (b) No (c) Cannot say	Emotional
6. The premier of France is: (a) LeFond (b) Debre (c) Delacroix (d) Saigon	Current events
7. U Thant is a native of: (a) Thailand (b) Burma (c) Siam (d) South Vietnam	Current events
8. I prefer bus rides to subway rides. (a) Yes (b) No (c) Cannot say	Neutral
9. Most people have to settle for a marriage partner very different from their dreams. (a) Yes (b) No (c) Cannot say	Emotional
10. Final grades should not be based so much on test marks. (a) Yes (b) No (c) Cannot say	Test-anxiety
11. If it takes three men two days to lay a 2-2/3 inch pipe, how long will it take two men? (a) 1 1/3 days (b) 3 days (c) 4 days (d) 2 2/3 days	Mathematics
12. Spring is the most pleasant time of the year. (a) Yes (b) No (c) Cannot say	Neutral
13. Most mothers are entirely too domineering. (a) Yes (b) No (c) Cannot say	Emotional
14. The square root of 563.12 is: (a) 2.373 (b) 23.73 (c) .237 (d) 70.32	Mathematics
15. The value of $X$ in the following equation is: $2X = 50 - (6 \times 2)^2 + (303 \div 3)$ (a) 129 (b) 64.5 (c) 7.0 (d) 3.5	Mathematics

questions on current events, (TA) test-anxiety questions, (M) mathematics questions, and (E) emotional questions.

As a precaution against confounding "primacy" and stimulus value, a "neutral" question was always presented as the first stimulus in a total series. For the other 14 questions, cards were shuffled to determine the order of presentation. The same sequence was used for all *Ss*.

While instructions were being issued, *S* was "hooked up" to the psychogalvanometer. After a fairly stable basal skin resistance had been recorded, the first question was presented. By nodding his head, *E* signalled to his as-

sistant when each card was to be presented. The assistant sat before *S* with the 15 cards face down on his lap and turned one over at each signal. As *S* answered each question, his verbal reply was recorded by *E*.

The GSR response to each question was complex. It usually consisted of a response to the question as well as a response to the answer, and the time for the total combined response varied. *E* signalled the onset of the next question when, from the ammeter recording, he decided that a base was again achieved.

After the 15th question had been presented, and *S* had made his response, *S* was queried on the taking of medication prior to the experiment to eliminate the possible effect of this variable. *Ss* were enjoined to secrecy regarding the specific questions used in the experiment, but were permitted to describe the general setup to classmates.

### C. RESULTS

Responses to stimuli were recorded as the highest galvanometer reading between the recorded onset of the stimulus (presentation of question) and the completion of the response. The rectifier was calibrated with a known decrease in ohms per space traversed by the marker pen, permitting a record of the response in absolute ohms decrease with a mean error, probably, of 50 ohms.

Because of a Spearman rank-difference correlation ( $\rho$ ) of  $-.052$  between each *S*'s rank on the basal skin resistance and his absolute response rank, the authors conclude that response is not a function of base. At this point should be mentioned the careful cleansing of the palmar area to which the electrodes were applied, and also the friction applied to this area with a tongue depressor, and the 20-minute application of the electrode jelly before recording. These procedures should have had the effect of lowering basal skin resistance and of stabilizing the chemical effects of the electrode jelly. Edelberg (5) notes that various detergents and conducting media differ in their effects, but have the general effect of decreasing basal skin resistance and decreasing GSR.

From Table 2, it can be seen that response to the first question, a neutral question, is greater than response to any other question. It can also be seen that the first question of each type yields a greater response than do subsequent questions of the same type. For example, the response to the first mathematics question, No. 11, is greater than the response to Questions 14 and 15.

Table 3 presents the results of an analysis of variance. Variables were (*a*) order of stimulus presentation, (*b*) type of stimulus, and (*c*) individuals.

The *F* test corroborates the idea that the most important variable is the

TABLE 2  
MEAN RESPONSE IN OHMS

Sequence by type	Neutral	No.	Current Events	Question type and identifying number						Mathematics	No.
				No.	Test-anxiety	No.	Emotional	No.			
1st	1,485	1	1,036	2	789	3	1,155	5		1,069	11
2nd	475	8	723	6	723	4	630	9		663	14
3rd	449	12	604	7	525	10	465	13		861	15
Mean	803		787		679		750			864	

TABLE 3  
ANALYSIS OF VARIANCE FOR GSR RESPONSES TO 15 STATEMENTS BY 25 SUBJECTS

Source	<i>SS</i>	<i>df</i>	<i>F</i>
Stimulus (A)	51.909	4	1.514
Order (B)	758.836	2	24.477**
Subjects (C)	4897.610	24	23.815**
A × B	322.291	8	4.822*
A × C	822.591	96	—
B × C	744.030	48	—
A × B × C	1604.009	192	—
Total	9201.276	374	

\*  $p \leq .01$ .

\*\*  $p \leq .0001$ .

order of stimulus presentation ( $F = 24.477$ ), but variations among individuals also prove significant ( $F = 23.815$ ) and interindividual differences are significantly greater than intraindividual differences. Type of stimulus, however, yields no significant differences.

#### D. DISCUSSION

According to our data, the presentation of any stimulus for the first time is not only critical in evoking GSR but, in many instances, may account for the major portion of a response. The orienting reflex, which Robinson and Gantt (13) call the questioning reaction, apparently is a composite of response to surprise, novelty, and primacy. The first time there is anything different about a stimulus—whether change of color or tone or type of question—elicits alertness via the orienting or questioning response. In modalities other than the GSR, this phenomenon has been observed in cardiac, respiratory and motor behavior.

No matter how well prepared *S* may be, and no matter how similar stimuli are, the *first of anything* will evoke a good portion of the GSR, if *S* is responding at all.

The stimuli used in this study were similar in that all were questions printed on the same kinds of cards and were presented under the same circumstances. They differed in content (type), and the authors expected much greater type-differences than were found. Even questions that concerned attitude toward tests were not particularly evocative. After initial presentation, responses to all stimuli fall off rapidly.

In 1922 Smith (15), and in 1928 Jones and Wechsler (6), found that, in word-association tests, emotion-provoking words produce the largest GSRs. In 1920 Prideaux (10), and in 1928 Wechsler (16), reported an increase in skin conductance during mental work. This increase, Wechsler



and Prideaux believed, results not from the mental work itself, but from the fear of giving wrong answers. In the present study, which used emotionally toned questions as well as questions that involved mental work, no difference in the mean GSR of these two types of questions was found. Neither did these two mean GSRs differ significantly from the mean GSR of neutral questions.

In conditioning experiments, the most substantial evidence of conditioning is often the response to the first presentation of the CS without the US. If it is *firstness* to which *S* largely responds, conditioning studies might be more valid if trials other than the first trial without the US are considered.

### REFERENCES

1. BERLYNE, D. E. *Conflict Arousal and Curiosity*. New York: McGraw-Hill, 1960.
2. ———. Conflict and the orienting reaction. *J. Exper. Psychol.*, 1961, **62**, 476-483.
3. COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *J. Exper. Psychol.*, 1938, **22**, 244-268.
4. DAVIS, R. C. Factors affecting the galvanic reflex. *Arch. Psychol.* (N.Y.), 1930, **18**, 5-64.
5. EDELBERG, R., GREINER, T., & BURCH, N. R. Some membrane properties of the effector in the GSR. *J. Appl. Physiol.*, 1960, **15**, 691-696.
6. JONES, H. E., & WECHSLER, D. Galvanometric technique in studies of association. *Amer. J. Psychol.*, 1928, **40**, 607-612.
7. LANDIS, C., & HUNT, W. A. The conscious correlates of the galvanic skin response. *J. Exper. Psychol.*, 1935, **18**, 505-529.
8. PATTERSON, E. A qualitative and quantitative study of the emotion of surprise. *Psychol. Monog.*, 1930, **40**, 85-108.
9. PAVLOV, I. P. *Lectures on Conditioned Reflexes*. New York: International Univ. Press, 1928.
10. PRIDEAUX, E. The psychogalvanic reflex, a review. *Brain*, 1920, **43**, 50-73.
11. PROSSER, C. L., & HUNTER, W. S. The extinction of startle responses and spinal reflexes in the white rat. *Amer. J. Physiol.*, 1936, **117**, 609-618.
12. RAZRAN, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81-147.
13. ROBINSON, J., & GANTT, W. H. The orienting reflex (questioning reaction): Cardiac, respiratory, salivary and motor components. *Bull. Johns Hopkins Hosp.*, 1947, **80**, 231-253.
14. SEWARD, J. P., & SEWARD, G. H. The effect of repetition on reaction to electric shock. *Arch. Psychol.* (N.Y.), 1934, **25**, 7-103.
15. SMITH, W. *The Measurement of Emotion*. London: Kegan Paul, 1922.
16. WECHSLER, D. The measurement of emotional reaction. *Arch. Psychol.* (N.Y.), 1925, **12**, 5-181.

Department of Psychology  
Hunter College  
New York 21, New York

## PREDICTING ORGANIZATION IN CHANGED COMMUNICATION NETWORKS: III.\*<sup>1</sup>

*Department of Sociology and Anthropology, Emory University*

---

ARTHUR M. COHEN

---

### A. INTRODUCTION

#### 1. *Previous Research*

This study is part of a series of laboratory experiments aimed at understanding how groups organize for work when their organizational structures, defined by communication networks,<sup>2</sup> have been changed.<sup>3</sup> In the absence of prior experience in a different network and without external direction, a group tends to develop a problem-solving system that is typical for and particular to the type of network in which its members try to solve problems. The typical system generally conforms to the outline of the network. For example, in a network that has no communication restrictions the typical system tends to be decentralized and role-undifferentiated. The typical system, however, may be less efficient for a particular task than one or more others that can possibly be developed within the network. Groups whose networks have been changed are more aware than nonchange groups of the possibility of organizing their activities alternatively and atypically because of their prior experience in a different network with a different typical problem-solving system. They are also more skilled in the actual operation of one alternative, their former problem-solving system. A group whose communication network has been changed might develop a problem-solving system in the subsequent network that would be more, or less, efficient than the typical one, depending on the task, by making adaptations of its former system. Studies have been conducted involving changes from Wheel to Circle networks (3, 6), Wheel to Completely Connected networks (4)

---

\* Received in the Editorial Office on February 27, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This report is based on research sponsored by the Office of Naval Research Group Psychology Branch, under Contract Nonr-1141, Project Nr 177-389. The research was conducted at Case Institute of Technology.

<sup>2</sup> A communication network is a physical and formal structure that defines who can communicate directly with whom. It sets the limits on how members can interact. It does not specifically program interactions, because even within very limiting structural constraints members can develop more than one operational pattern or problem-solving system.

<sup>3</sup> For a summary of the studies see Cohen (1, 2) and Cohen and Bennis (4).

and Chain to Completely Connected networks (2), in order to determine the conditions under which a group will make use of its former system. The results of these studies support the following theorization: "The problem-solving system used in an antecedent network will be adapted for use in a subsequent network when, on the basis of the dominant principle of organization underlying it (i.e., centralization), a system could be established in the subsequent network that would be more efficient than its typical system." The three combinations of networks were used because they produced special sets of conditions that represented the three categories of single-network change whose tests were needed to evaluate the usefulness of the above theorization. Each category was distinguished from the others on the basis of the degree of adaptation that could maximally occur in the subsequent network (because of its structure) relative to the degree to which the prior system employed the organizational principle. In this regard, adaptation can be considered maximum for a group whose subsequent problem-solving system uses the organizational principle underlying the former system to a smaller degree (category A), to the same degree (category B), or to a greater degree (category C). Maximum adaptation would be represented in category A by the establishment of a relay system following the change from a Wheel to a Circle network, in category B by the establishment of a central-hub system following the change from a Wheel to a Completely Connected network, and in category C by the establishment of a central-hub system following the change from a Chain to a Completely Connected network. Maximum adaptation occurred in each of the above network changes in accordance with the theorization.<sup>4</sup>

## 2. *Aim*

These results, however, do not provide direct support for the theorization when it is expressed negatively: "The problem-solving system used in an antecedent network will *not* be adapted for use in a subsequent network when, on the basis of the dominant principle of organization underlying it (i.e., centralization), a system could be established in the subsequent network that

---

<sup>4</sup> In each of these studies the dominant organizational principle underlying the problem-solving systems of the antecedent networks was centralization of information-collection and answer-distribution. Based on centralization, maximum adaptation was equivalent to the most centralized system that could possibly have developed in each of the subsequent networks. Operationally, the most centralized system within a network has the smallest number of steps between the central member and the member most distant from him. The most centralized systems possible were more efficient than the typical systems of the subsequent networks for the tasks used in these studies.

would be *less* efficient than its typical system." In contrast to its positive form, which was intended to answer the question, "Under what conditions will a group make use of its former system?" the negative form of the theorization is intended to answer the question, "Under what conditions will a group *not* make use of its former system?" Data that can be used in direct support or nonsupport of the negative form of the theorization have to come from an experiment whose conditions correspond with those specified in its expression. This study was intended to provide such data.

## B. HYPOTHESIS

It was predicted that groups which had undergone change from a Wheel to a Completely Connected network would reject the use of their former centralized problem-solving system in favor of the decentralized system that is typical for the Completely Connected network. This prediction was derived from the negative form of the theorization. The change from a Wheel to a Completely Connected network and the special task that was used produced conditions that corresponded with those specified in the negative form of the theorization: (a) For the task that was used, the typical centralized system of the Wheel is less efficient (using times taken to solve problems as the criterion) than the typical decentralized system of the Completely Connected network. (b) The Completely Connected network permits the establishment of any system. (c) Using the dominant principle of organization—centralization—underlying the former system would produce a centralized system in the subsequent network that would be less efficient than its typical decentralized system.

## C. METHOD

### 1. Apparatus

Subjects were seated around a circular table so that each was separated by a vertical partition extending from the center to a foot beyond the edge of the table. At the center of the table was a five-layer pentagonal box with slots (openings) permitting subjects to push written messages to those with whom they were allowed to communicate. This apparatus allowed for the arrangement of any kind of communication network by blocking the appropriate slots with cardboard templates.

### 2. Procedure

Fifty subjects, paid male volunteers drawn from undergraduate classes, were randomly assigned to 10 groups of five men each. Subjects in each



group were randomly assigned to positions (which were distinguished by different colors) in the networks and were used for only one Wheel-to-Completely-Connected network sequence. Each group spent about three hours in the experimental room. No subjects were used who had any knowledge of the experiment or who were color-blind. Subjects were not informed of the kind of network in which they worked, but could find out for themselves. By sending messages about who was communicating to whom, subjects were able to find out what particular functions were being performed by whom. Before solving problems, subjects read instructions that stressed the importance of working efficiently and quickly.

Each group tried to solve 35 problems (trials). Only written communication was allowed among those who had open channels. There were no restrictions on the kind or amount of information in a message. A trial was ended when all five members of a group had registered answers by depressing switches, representing possible answers, that were contained in a switch box in each booth. Change from the Wheel was introduced after 15 trials, after stabilization of both problem-solving system and time had taken place (see Tables 1 and 2). The physical changes in networks were accomplished by substituting Completely Connected network templates for the Wheel templates. Changes were thus introduced without requiring any physical movements on the part of the subjects.

FIGURE 1  
EXAMPLE OF THE TASK

---

Trial No. 36  
Your Color: Blue

---

Your Story: ⑤ (also the number of your missing part)

- (a) A traveler about to set off on a journey saw his dog standing at the door.
  - (b) He said sharply: "What do you stand there gaping for? Everything is ready but you."
  - (c) "O, master! I am quite ready; it is you for whom I am waiting."
- 

- ④ (the number of someone else's missing part)  
*Then he shook down the acorns.*
- 

### 3. Task

The task was specifically designed so that a centralized problem-solving system would be less efficient than a decentralized one. Five additional groups were run in only the Completely Connected network. Their times were compared with those in the Wheel network for the 10 change groups. It was determined for this task that groups in the Completely Connected networks



TABLE 1  
NUMBER OF TRIALS TO STABILIZATION OF CENTRALIZED PROBLEM-SOLVING SYSTEMS IN WHEEL NETWORKS

Group No.	$W-CC_1$	$W-CC_2$	$W-CC_3$	$W-CC_4$	$W-CC_5$	$W-CC_6$	$W-CC_7$	$W-CC_8$	$W-CC_9$	$W-CC_{10}$
Stabilization by trial number	2	4	5	3	6	8	4	3	3	4

TABLE 2  
ANALYSIS OF VARIANCE OF TRIAL TIMES IN WHEEL NETWORKS  
FOR TRIALS 11 THROUGH 15

Source of variance	Sum of squares	df	Mean square	F
Between trials	2,310.92	4	577.73	< 1
Between groups*	75,553.12	9		
Interaction:** groups $\times$ trials	36,147.48	36	1,004.10	

\* Bartlett's Test of Homogeneity of Variance had previously shown that the variances of these groups were homogeneous.

\*\* This was used as the error term for Between Trials.

developed decentralized problem-solving systems that were more efficient (using times taken to solve problems as the criterion)<sup>5</sup> than the centralized problem-solving systems developed by groups in the Wheel networks (see Table 3).

TABLE 3  
ANALYSIS OF VARIANCE OF TIMES\* IN WHEEL AND COMPLETELY CONNECTED NETWORKS

Source of variance	Sum of squares	df	Mean square	F
Between networks**	110,324.22	1	110,324.22	17.99
Between groups in the same network***	80,638.26	13	6,202.99	(W > CC)
Between trials	2,041.81	4	510.45	< 1
Interaction: trials $\times$ networks	615.91	4	153.98	< 1
Residual****	41,527.48	52	798.61	

\* Time scores were for the last five trials in each network: for trials 11 through 15 in the Wheel network and for trials 16 through 20 in the Completely Connected network. In both networks, stabilization of problem-solving systems had taken place. Times were in seconds.

\*\* Additional analysis, not shown here, revealed that Com-con groups took significantly shorter times to solve problems than Wheel groups, when the five fastest trials in each network were compared.

\*\*\* This was used as the error term for Between Networks. Bartlett's Test of Homogeneity of Variance had revealed homogeneous variance among groups in the Wheel network and among groups in the Com-con network.

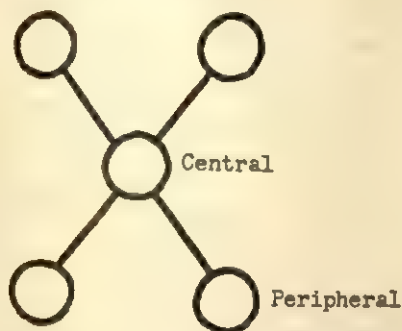
\*\*\*\* This was used as the error term for Between Trials and the Interaction: Trials  $\times$  Networks.

The task was as follows. On each trial, each subject was given a card (see Figure 1) which had the following information on it: (a) the trial number; (b) the color of the booth identifying the subject: red, white, yellow, green, or blue (which remained constant throughout the trials); (c) three parts, designated as a, b, and c, of a four-part story belonging to him; (d) one part

<sup>5</sup> Groups in Completely Connected networks did not significantly differ in number of correct trials from those in Wheel networks. Few trials were incorrect in either condition.

of a story belonging to another member. His own story was referred to by number 1, 2, 3, 4, or 5. The part that he had which belonged to someone else's story was identified by a different number from 1 to 5 and corresponded with the story number of the other member to whom the extra part belonged. Thus, on each trial, each member had a card on which there were three parts of a four-part story and one part that belonged to some unknown other member's story. A trial was ended when each member registered answers (by depressing appropriately identified switches) denoting 1) the color identifying the person who held the missing part for his story on that trial and 2) the location of the missing part in the story: between parts *a* and *b* or between parts *b* and *c*. Subjects were correctly instructed that their stories would vary from trial to trial and that the person holding their missing parts might change from trial to trial. Consequently, the color that each would need to register as part of the answer might need to be changed correspondingly. They were also informed that the numbers designating their stories might change from trial to trial. Subjects were instructed that a trial would not be

WHEEL NETWORK



COMPLETELY CONNECTED NETWORK

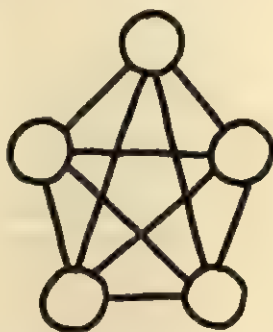


FIGURE 2  
COMMUNICATION NETWORKS AND THEIR CHARACTERISTICS

ended until every member had registered a two-part answer. The stories that were used were modified versions of Aesop's Fables. They were equated for both length of total story and for length of component parts. Stories were randomly assigned to trial numbers.

In the Wheel network (see Figure 2) each of four members can communicate to the fifth person, but to no one else. The fifth person can communicate to everyone directly. The others have to go through him if they want to com-

municate to someone else. These four members are equal to each other in the communications restrictions imposed on them, but are much more restricted than the fifth member, the occupant of the central position. For the task used in this study, members in the Wheel network typically used a centralized problem-solving system in which each of the four peripheral members sent to the occupant of the central position the number of the story that he had, the number of the part that he had that belonged to someone else, and a full quotation of this part. The central person then matched those who held and needed each missing part; he copied each missing part, designating the color of the person who held it; and he sent such information to the appropriate peripheral member. Each peripheral member located the needed part in his story sequence and registered both the color of the person who originally had held the part and the location (between *a* and *b* or between *b* and *c*) in his story.

In the Completely Connected network (Figure 2) every member has equal communication opportunities. This network imposes no communication restrictions on members. Consequently, any problem-solving system can be developed, including the kind that typically develops in the Wheel network. For the task used in this study, groups working in just the Completely Connected network typically used a decentralized problem-solving system. Each member received his needed missing part directly from the member who had it. This was typically preceded by either one or the other of the following procedures: either (*a*) each member directly located who had his missing part by sending out messages containing the number of the story part he needed to the other four members, or (*b*) a member found out from another member who had his missing part. Both procedures were decentralized in that the copying and the processing of missing parts were equally and directly shared by members.

#### D. RESULTS

How groups organized for solving problems (who communicated what to whom) was determined by content analysis of message cards. A unit was defined as a sentence or a meaningful part of one. Because of the nature of the task, communication channels for each group for each trial were analyzed separately and with particular emphasis on message contents comprising what members had and needed. The coded messages for each trial were placed on a grid (see Table 4).

##### 1. *Content Analysis*

The following notations were used: N = need, H = have, Q = quotation of a part belonging to another member, and i = a question. Subscripted num-

bers 1, 2, 3, 4, or 5 refer to the story that a member had or to the missing part. Superscripted letters W, Y, B, R, or G refer to subjects. Several examples are as follows:

- $N_3H_4Q_4$  = I need the part for story #3. I have the part for story #4.  
 It is—  
 $H_8^1$  = Who has the part to story #3?  
 $H_8^{B1}$  = Does Blue have the part to story #3?  
 $N_8^{R1}$  = Does Red need the part to story #3?  
 $N^{B1}$  = What # does Blue need?  
 $Q_8^1$  = What is the quotation of the part to story #3?  
 $\overline{H_8}$  = I don't have the part to story #3.

Following the content analysis, the last five trials of each group were evaluated in terms of whether or not the system was the same for each of these trials and what the system was. Table 4 represents the ways in which a centralized and a decentralized system would appear, following content analysis, on any one trial.

TABLE 4  
PROBLEM-SOLVING SYSTEMS

Sender	R	B	Receiver Y	G	W
<i>Centralized system</i>					
R					$N_1H_5Q_5$
B					$N_2H_8Q_3$
Y					$N_4H_1Q_1$
G					$N_5H_2Q_2$
W	$H_1^YQ_1$	$H_2^GQ_2$	$H_4^WQ_4$	$H_5^BQ_5$	
<i>Decentralized system</i>					
R		$H_3N_1$	$H_3N_1Q_3$	$H_3N_1$	$H_3N_1$
B	$H_4N_2$		$H_4N_2$	$H_4N_2Q_4$	$H_4N_2$
Y	$H_5N_3$	$H_5N_3$		$H_5N_3$	$H_5N_3Q_5$
G	$H_1N_4Q_1$	$H_1N_4$	$H_1N_4$		$H_1N_4$
W	$H_2N_5$	$H_2N_5Q_2$	$H_2N_5$	$H_2N_5$	

## 2. Problem-Solving Systems

The test of the prediction was based on the comparison between the groups observed in this experiment, designated as W-CCb, and those groups in a prior experiment (4), designated as W-CCa. W-CCa groups had used a task (finding common symbols) that was more efficiently solved in a centralized system. The W-CCa groups developed centralized problem-solving systems in Completely Connected networks that had been anteceded by Wheel networks.



In contrast, the W-CCb groups in the present study developed significantly more decentralized problem-solving systems (Table 5). Of the 10 W-CCa groups, one developed a decentralized problem-solving system. Of the 10 W-CCb groups, seven developed decentralized problem-solving systems. This latter result does not differ significantly from what was found to have been

TABLE 5  
FISHER'S EXACT TEST OF PROBLEM-SOLVING SYSTEMS IN W-CCA AND W-CCb NETWORKS

Network	Centralized	System	Decentralized
W-CCa	9		1
W-CCb	3		7

$p = .01.$

typical for groups that solve problems only in the Completely Connected network: (a) of five control groups that used the same task as in this study, all developed decentralized systems; (b) of 10 groups using the common symbol task, eight developed decentralized systems and two developed centralized systems (4).

### E. DISCUSSION

The results of this study provided support for the following theorization: "The problem-solving system used in an antecedent network will *not* be adapted for use in a subsequent network, when, on the basis of the dominant principle of organization underlying it (i.e., centralization), a system could be established in the subsequent network that would be *less* efficient than its typical system."

The design of this study provided a necessary complement to the designs of the three studies that preceded it (2, 4, 6). From these designs it could not be determined unambiguously whether the use of prior experience was based on habit or insightful application of problem-solving principles. Behavior is considered based on insightful application, in contrast to habit, if selections from available channels stem from assessment of the efficiency of the prior system, or the maximum possible adaptation of it, relative to the efficiency of the typical system of the subsequent network. These selections are expected to be cumulative until only those channels remain used in specialized ways that define the more efficient problem-solving system possible within the limits of the subsequent network and within the limits of experience with the typical problem-solving systems of the antecedent and the subsequent networks. In the interest of efficiency, a group whose network has been changed might find it necessary to reject the use of the former system or the

maximum possible adaptation of it in favor of the typical system of the subsequent network. Habitual behavior, in contrast to behavior based on insightful application, would be exemplified by the continued use of the former system, if possible, or the maximum possible adaptation of it.

The design of this study involved change from a Wheel to a Completely Connected network. A task was used for which the typical, centralized system of the Wheel network was less efficient than the typical, decentralized system of the Completely Connected network. In this circumstance, predictions based on the explanations of habit and insightful application would have led to mutually contradictory results. Based on the explanation of habit, groups would be expected to have continued using their former centralized systems in the subsequent networks. Based on the explanation of insightful application, groups would be expected to have rejected the use of their former centralized systems in favor of decentralized systems typical for the subsequent networks. The results support insightful application as the more valid explanation. The conditions in which groups underwent change in this experiment provided an acceptable test of habit and insightful application as possible explanations for the results of previous research. Subjects develop a strong set in the Wheel network. When it is changed to a Completely Connected network, there are no communication obstacles that require changes in role relationships and that might interfere with a group's direct and habit-based transfer of its accustomed centralized system. However, in order for a group to develop a decentralized problem-solving system, it has to achieve a considerable amount of agreement by group members to dispense with centralized copying and processing of data and to assume new role relationships involving direct transmission of relevant information from one member to another. The tendency to persist in the use of previous, less efficient procedures provided vigorous, but generally and ultimately unsuccessful, competition with attempts by group members to establish different problem-solving systems. In each of the groups that decentralized, there was a transition period following the network change in which proposals for continuation of the former centralized system competed with proposals for decentralization.

Special note needs to be made of the nature of the task and its interaction with problem-solving systems. In the course of trying to develop a task for which a decentralized problem-solving system would be more efficient than a centralized system, a number of potential tasks were investigated but not used; e.g., the mottled marble task by Smith (10) and arithmetic problems similar to those used by Shaw (9) and by Mulder (8). These and other tasks were not used because preliminary work with them had led to the ex-

pectation that a centralized system would be better than a decentralized system once its (the centralized system's) operation had become crystallized and standard forms for the identification of task parts and the determination and distribution of answers had been established.<sup>6</sup> The arithmetic problems used by Shaw and by others require only one solution for all members on any given trial; they permit the development of standard forms for the identification of task parts and the determination and distribution of answers; and, when sufficient in number, they ensure the establishment of stable and fully operational problem-solving systems. When such conditions are present, the demands on the central member of a centralized system for copying and processing data are considerably reduced. Consequently, unlike the task used in this study, tasks like those mentioned above may be considered *less complex* and more like the simple and routine common-symbol problems that were found to have been more efficiently solved in centralized systems (4, 5, 6, 7).

The task used in this study was less efficiently solved in a centralized system even after its decision structure had fully developed and had reached stability. The reasons for this became evident after close observations of the responses of pilot group members performing the task and analyses of its logical and psychological requirements. For this task, members registered different answers on each trial. In a centralized system two kinds of activities characterized the role of the central person: (a) an integrating function (i.e., the matching of missing parts to their respective stories), and (b) a copying-processing function (i.e., the copying of each missing part on a separate message card and its transmission to an appropriate target). It was the latter demand on the central member that produced the significant dif-

<sup>6</sup> The following is an example of the arithmetic task used by Shaw and Mulder: A small company is moving from one office building to another. It must move four kinds of equipment: (a) chairs, (b) desks, (c) filing cabinets, and (d) typewriters. How many trucks are needed to make the move in one trip? The eight items of information needed to solve the problem are that the company owns a total of 12 desks, 48 chairs, 12 typewriters, 15 filing cabinets, and that one truckload can take 12 typewriters, or three desks, or five filing cabinets, or 24 chairs. The statement of the problem was typed completely on each of four separate cards, and each of the eight items of information was typed on a separate card. Each subject was given one of the problem cards; the information was equally divided over the four group members. For completion of the group task, each subject has to know the problem solution. This task can be routinized and standardized in the following way: let  $n$  = the number of items in each category that would fill one truck (i.e., 24 chairs); let  $N$  = the number of items in each category that have to be moved (i.e., 48 chairs); let  $T$  = the total number of trucks required to move all the equipment. Then,  $T = N1/n1 + N2/n2 + N3/n3 + N4/n4$ . Thus this problem can be reduced to more routinized and simple form similar to the common-symbol problem in which the central member collects dispersed data, places them in the formula for  $T$ , and distributes the value of  $T$ .

ference in efficiency between the centralized and the decentralized problem-solving systems. Shaw (9) has hypothesized that "as complexity increases, the greater efficiency of a more centralized network decreases; ultimately becoming less efficient when the central person has become saturated; i.e., when his optimal output level is exceeded." Due to its solution requirements, the task used in this study produced the kind of saturation that Shaw had hypothesized. It led to the greater efficiency of the decentralized system, which he had predicted on the basis of saturation. Two conclusions concerning the inappropriateness and the appropriateness of the saturation hypothesis are drawn from the above discussion. It is *inappropriate* when: (a) integration rather than copying and processing of data is the major prerequisite of efficient solutions of the task; (b) on a given trial, the solution is the same for all members; (c) a sufficient number of problems will have been solved by a group to ensure that a centralized system has been fully developed and has reached stability; and (d) the problems from trial to trial, while requiring different solutions, are of the same categorical type, so that standard forms for the identification of task parts and the determination and the distribution of answers can be established, thus reducing the demands on the central member in a centralized system. It is *appropriate* for explaining the superiority of decentralized over centralized problem-solving systems for tasks, such as the one used in this experiment: (a) in which members register different answers on a given trial, (b) in which a major proportion of the problem-solving demands on members involves extensive copying and processing of data, and (c) in which the nature of such demands is not significantly altered by the development of standard forms for the identification of task parts and the determination and the distribution of answers.

#### F. SUMMARY

This study was intended to test the following hypothesis: When groups undergo change from a Wheel to a Completely Connected network, using a task for which a centralized problem-solving system is more efficient than a decentralized problem-solving system, they will reject the use of the centralized system followed in the antecedent Wheel network in favor of the decentralized system typical for the subsequent network. This hypothesis was derived from the following theorization: "The problem-solving system used in an antecedent network will *not* be adapted for use in a subsequent network when, on the basis of the dominant principle of organization underlying it (i.e., centralization), a system could be established in the subsequent network that would be *less* efficient than its typical system."



Ten groups of five men each underwent change from a Wheel in which they solved 15 problems, to a Completely Connected network in which they solved 20 problems. Subjects worked on problems for which the typical centralized system of the Wheel was less efficient than the typical decentralized system of the Completely Connected network. Each of these problems required the location of the missing part to a different story possessed by each subject, the identification of the color representing the other member who held the part, and appropriate placement of the part in a story sequence. Problem-solving systems were measured by analyzing the content and direction of flow of messages sent by members.

The behaviors of these change groups, designated as W-CCb, were compared with those of similar groups, designated as W-CCa, which had also undergone change from a Wheel to a Completely Connected network, but which had used a task for which a centralized system was more efficient than a decentralized system. The results showed that W-CCb groups developed significantly more decentralized systems than W-CCa groups and in a proportion that was not significantly different from what was typical for nonchange control groups in the Completely Connected network. These results confirmed the hypothesis and supported the theorization, based on previous studies, which was intended to answer the question: Under what conditions will a group not make use of its former problem-solving system? The design of this study made possible an unambiguous evaluation of habit and insightful application as alternative explanations for the use of prior problem-solving experience. Insightful application was clearly supported.

Two conclusions were reached following an analysis of the nature of the task and its interaction with problem-solving systems. A *centralized system is more efficient* when: (a) integration rather than copying and processing of data is the major prerequisite of task efficiency; (b) on a given trial, the solution is the same for all members; (c) a sufficient number of problems will have been solved by a group to ensure that a centralized system has been fully developed and has reached stability; and (d) the problems from trial to trial, while requiring different solutions, are of the same categorical type, so that standard forms for the identification of task parts and the determination and the distribution of answers can be established, thus reducing the demands on the central member in a centralized system. A *decentralized system is more efficient* than a centralized system for tasks, such as the one used in this study: (a) in which members register different answers on any given trial, (b) in which a major proportion of the problem-solving demands on members involves extensive copying and processing of data, and (c) in which the nature of



such demands is not significantly altered by the development of standard forms for the identification of task parts and the determination and the distribution of answers.

## REFERENCES

1. COHEN, A. M. Changing small-group communications networks. *Admin. Sci. Quart.*, 1962, **6**, No. 4.
2. ———. Predicting organization in changed communication networks: II. *J. of Psychol.*, 1964, **57**, 475-499.
3. COHEN, A. M., & BENNIS, W. Continuity of leadership in communication networks. *Hum. Rel.*, 1961, **14**, 351-367.
4. ———. Predicting organization in changed communication networks: I. *J. of Psychol.*, 1962, **54**, 391-416.
5. COHEN, A. M., BENNIS, W., & WOLKON, G. The effects of continued practice on the behaviors of problem-solving groups. *Sociometry*, 1961, **24**, 416-431.
6. ———. The effects of changes in communication networks on the behaviors of problem-solving groups. *Sociometry*, 1962, **25**, 177-196.
7. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. Abn. & Soc. Psychol.*, 1951, **46**, 38-50.
8. MULDER, M. Communication structure, decision structure and group performance. *Sociometry*, 1960, **23**, 1-14.
9. SHAW, M. E. Some effects of problem complexity upon problem solution efficiency in different communication networks. *J. Exper. Psychol.*, 1954, **48**, 211-217.
10. SMITH, S. Communication pattern and adaptability of task-oriented groups. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1950.

*Department of Sociology and Anthropology*  
*Emory University*  
*Atlanta 22, Georgia*

## EVIDENCE OF RIGHT-LEFT DISCRIMINATION IN PLANARIANS\*<sup>1</sup>

*Department of Biophysics, Michigan State University*

---

W. C. CORNING

---

### A. INTRODUCTION

Following the demonstration by Thompson and McConnell (17) of classical conditioning in planarians with light (conditioned stimulus) and shock (unconditioned stimulus), various investigations have indicated that animals of this class are capable of learning in other types of situations (2, 6, 7, 10, 13). Most convincing are those studies that have shown that planarians can master a maze requiring them to discriminate between two stimuli in order to receive reinforcement. To date, successful discriminations have been made between a white and a black maze arm (10), a rough and a smooth maze floor (10), and a lighted and darkened chamber (2). It is even reported that some animals are capable of making a "two-bit" decision in the maze; i.e., subjects were trained to enter a darkened chamber if the floor of the maze was smooth and a lighted chamber if the floor of the maze was rough (1, 2).

The use of the two-choice maze to examine the learning capabilities of planarians obviates the possibility of shock sensitization and consequent pseudoconditioning to light that sometimes occurs in classical conditioning (4, 8, 9). The maze would be, therefore, a less equivocal learning situation in which to repeat some of the regeneration and cannibalism effects that have suggested a possible biochemical basis of memory in these animals (3, 14, 15, 18, 19). These studies have relied for the most part upon the light-shock classical conditioning paradigm, but the basic phenomena of ribonuclease destruction of memory (3) and the incorporation of a possible memory code via ribonucleic acid (14, 19) ought to be observable with a different kind of learned response.

The present investigation is the first of a series, the general purpose of which is to explore the maze-learning characteristics and capacities of the planarian.

---

\* Received in the Editorial Office on February 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The present paper is a portion of the research carried out under Grant NSG 475, National Aeronautics and Space Administration.

It is hoped that by a thorough study of the factors that are critical in this type of learning situation, a procedure will be developed that is optimal for use in attempts to replicate the previously mentioned regeneration and cannibalism effects.

Little is known concerning the extent to which the planarian is capable of using proprioceptive cues in its adaptive behavior. This study has examined the ability of planarians to discriminate between a right and a left turn in a T-maze and has also obtained information concerning the planarian's response instability in a two-choice situation.

## B. METHOD

### 1. Subjects

The subject population consisted of 31 *Dugesia dorotocephala* and two *Dugesia tigrina*. Each of the subjects was kept in a separate bowl containing approximately 225 ml. of pond water. The subjects were fed beef liver twice weekly and had their water changed after each feeding. Between maze sessions, the subjects were kept in areas where only indirect room lighting impinged upon them.

### 2. Apparatus

The mazes were of a "T" design and were constructed of clear plexiglass (see Figure 1). A total of five mazes was used to permit randomization of the maze from trial to trial and to avoid the possibility of subjects utilizing cues that were peculiar to any one maze. During a trial, a maze was situated directly over the subject's bowl and positioned so that only the ends of the angled arms were beneath the water level.

### 3. Procedures

Two groups of subjects were used, one group being trained to enter a right or a left arm ( $N = 17$ ) and the other group being permitted to select either arm to escape the maze ( $N = 16$ ). Reinforcement for maze running in both groups consisted of the return to the home bowls via one of the maze arms.

The trials were conducted with the room lights off. To ensure that the subjects were not influenced by any external light cues, the mazes were situated inside a box with a flat black interior. Each subject received five nonconsecutive trials per day and usually was presented with four or five trial sessions per week. The intertrial interval varied from 6 to 12 minutes.

To begin a trial, a maze was selected, cleaned, positioned over the subject's bowl, and moistened with water. The subject was picked up with a soft

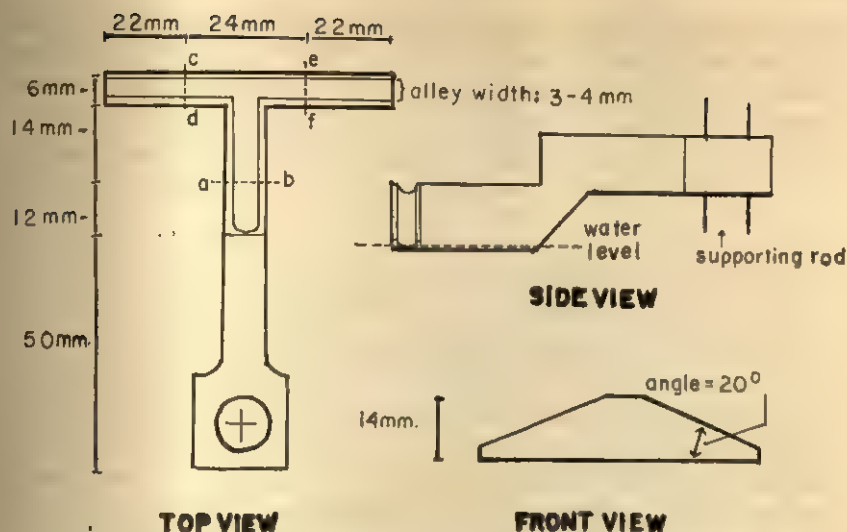


FIGURE 1  
T-MAZE DESIGN

[Maze design similar to that used by Humphries (10)]

camel's-hair brush and placed in the starting area of the maze. Care was taken to orient the subject's anterior end towards the choice point and to position the animal right side up. When line cd or ef (see Figure 1) was reached, the choice was recorded. After the subject had reached the end of an arm and entered its bowl, the maze was removed, cleaned, and another subject was presented its trial.

The first 10 trials for all subjects were considered preference-testing trials. The subjects of both groups were allowed to select either arm to return to their home bowl. Thereafter, the subjects being trained were not permitted to escape the maze via the arm of their preference during these first 10 trials, while the control subjects continued to have a free choice in the matter.

For the subjects being trained, self-correction procedures were used. When the subject made a wrong choice and began moving past lines cd or ef, it was given a gentle poke (with the brush) at its anterior end. This procedure usually elicited a withdrawal. These pokes were continued until the animal entered the correct arm. A correct response was recorded if the subject escaped the maze via the correct arm and did not receive a poke for a wrong choice. When a subject achieved a criterion of nine out of 10 correct turns, discrimination learning was assumed to have occurred.

In an attempt to quantify various aspects of the subjects' behavior in the maze, records of the following events were kept for each trial: (a) the number of pokes required at various points in the maze, (b) the number of times a subject turned around in the maze, (c) the number of withdrawals and incomplete choices at the choice point, and (d) the time it took a subject to traverse the maze from the starting line (ab) to edge of an arm.

## C. RESULTS

### 1. *Right-Left Discrimination*

The data obtained in the present experiment demonstrate that, while it is possible to establish a right-left discrimination in planarians, it is neither an easily acquired response nor, once attained, is it maintained over many succeeding trial sessions. A total of 10 out of the 17 subjects that received the training treatment were able to reach the response criterion of nine out of 10 correct choices, with the average number of trials to criterion equalling 35.8. In contrast, none of the 16 control subjects at any time reversed its original preference sufficiently to select the alternate side nine out of 10 times. The trained subjects did not, however, sustain their criterion level of performance, but would display a drop in the number of correct choices usually within 10 to 20 trials after the point of criterion achievement (some examples of this preference shift are presented in Figure 3A).

The average number of responses of the trained and the control groups over the 80 trials are shown in Figure 2. It can be seen that the mean performance of the trained subjects did not reach exceptionally high magnitudes in spite of the fact that 10 subjects had reached criterion. The response criterion was reached at different times (the number of trials to criterion ranged from 13 to 79) in these subjects and, because of their subsequent change of turning preference, the overall averages tended to be low. The means of the trained group departed in a statistically significant manner [ $t$  tests(5)] from the average response level during the preference tests at trial periods 10 ( $p < 0.01$ ), 20 ( $p < 0.02$ ), 30 ( $p < 0.02$ ), and 40 ( $p < 0.05$ ). The means of the control subjects at no point departed significantly from the performance level during the preference testing trials. The means of the trained and the control groups differed significantly at trial periods 30 ( $p < 0.01$ ) and 40 ( $p < 0.02$ ).

### 2. *Preference Instability*

The drop observed in the number of correct choices by the trained subjects was puzzling, and an examination of the individual response curves



served only to rule out such extraneous factors as feeding schedules, temperature, water quality, and time of day. The three subjects whose data are represented in Figure 3A began training on the same day. Consequently, they were always fed on the same days, they received trial sessions at the same

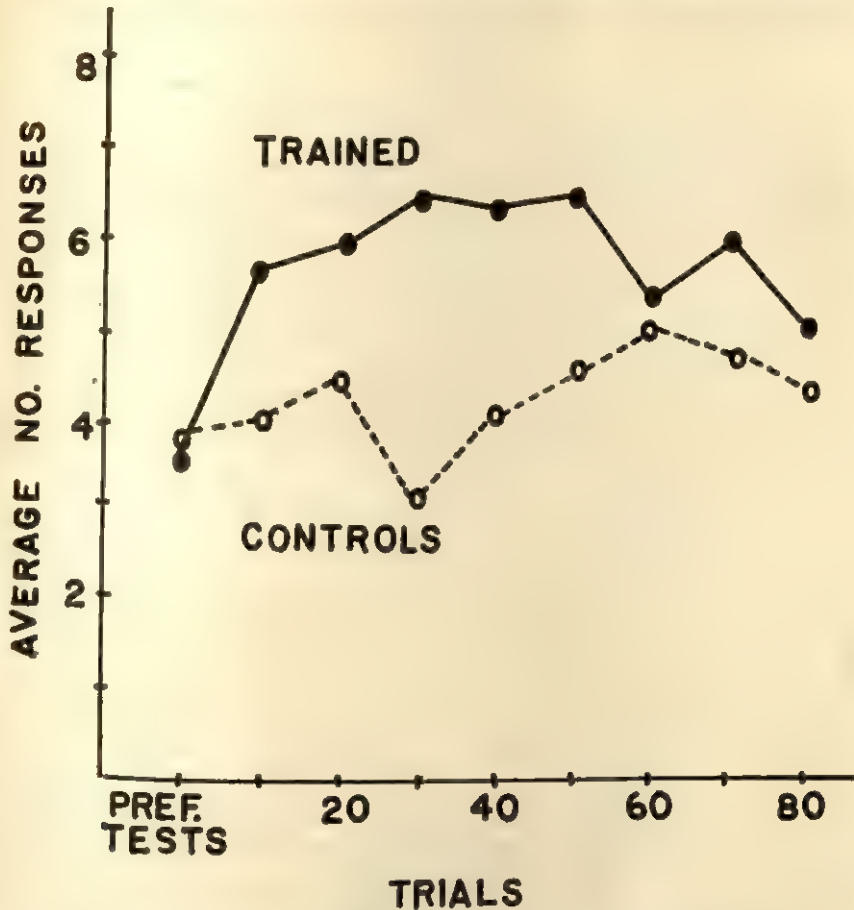


FIGURE 2

THE AVERAGE NUMBER OF RESPONSES OPPOSITE TO THE ORIGINAL PREFERENCE

time, and they were always at the same point in the training sequence. However, it can be seen that their response levels shifted at varying points over the trials. The response changes of three control animals that had also received concomitant trial treatments are displayed in Figure 3B. It can be observed

that these subjects also displayed a certain degree of response instability. In general, any long-term consistency in a turning preference was a rarity during the experiment, for the subjects of both groups.

In an attempt to determine if, perhaps, the addition of another cue might

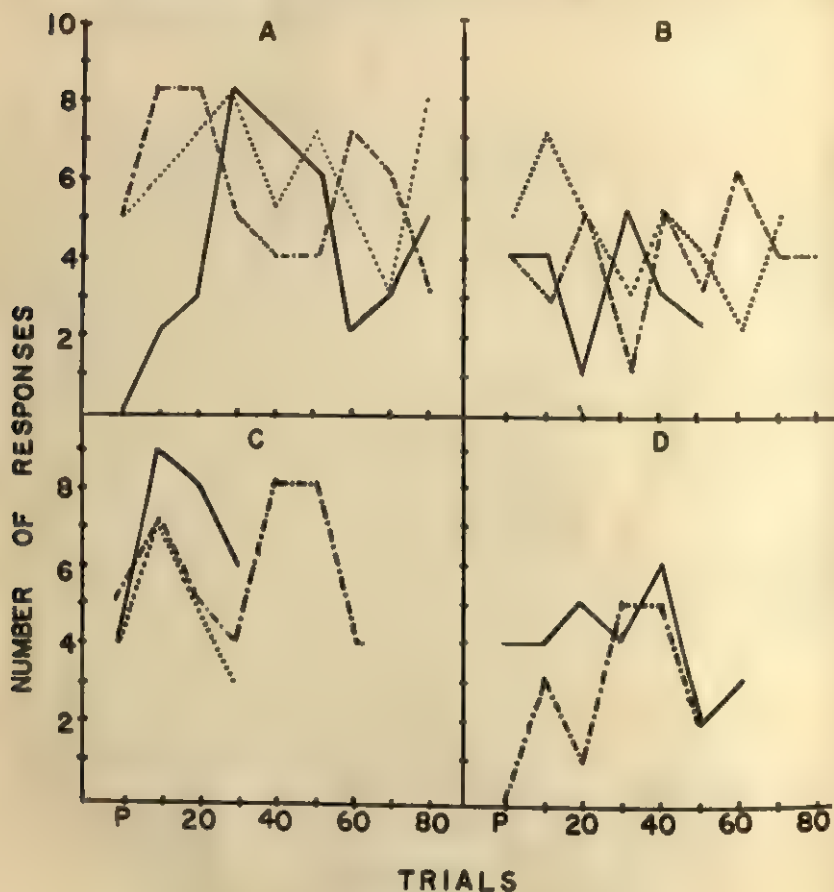


FIGURE 3

EXAMPLES OF INDIVIDUAL TURNING PREFERENCE CHANGES

A. Trained subjects that had received concomitant trial treatments. B. Control subjects that had received concomitant trial treatments. C. Responses of subjects that had been trained in a confounded cue (left-black *vs.* right-white) situation. D. Control subjects which had run in the confounded-cue situation. For both the trained and the control subjects, the responses plotted consist of the number of turns that were opposite to the turning choice exhibited during preference testing. The "P" on the abscissa represents the preference testing stage.

facilitate response stabilization, the right half of the box housing the maze was covered with flat white paneling and the left half was kept flat black. The ability of planarians to respond to differences in light flux is well known (11). As shown in Figure 3C and 3D, this modification has thus far failed to stabilize the subjects' response preferences.

### 3. Other Observations

A summary of the average number of pokes, latencies, and other measurements of the subjects' behavior is presented in Table 1. The results indicate that, in terms of the measurements that were made, the two groups did not differ over the 80 trials. An examination of the measurements at various points during the trials also failed to yield any consistent trend.

TABLE 1  
SUMMARY OF THE MEASUREMENTS MADE DURING TRIALS FOR THE TRAINED  
AND THE CONTROL GROUPS

Stimulus and response	Group		<i>p</i>
	Trained	Control	
Total pokes per subject	161.6	149.5	n.s.*
Total turns per subject	43.0	45.6	n.s.
Number of withdrawals and incomplete choices per subject	62.8	58.3	n.s.
Average latency per trial for each subject	62.2	61.8	n.s.

\* n.s. indicates that a *t* test failed to yield a significant difference between the means of the trained and the control groups.

### D. DISCUSSION

The ability of some planarians to learn a right-left discrimination adds to the growing repertoire of acquired behavior patterns reported for animals at this phylogenetic level [for a review, see Jacobson (12)]. However, the inability or unwillingness of the subjects to adhere to a response choice that permits them to escape the maze remains enigmatic. Best (2) has reported a similar phenomenon in a light-dark discrimination problem, but was able to improve upon response stabilization by supplying the subjects with a less-constricting goal chamber upon completion of a correct choice. In the present experiment, the goal for a subject consisted of its home bowl—presumably a positive chamber, because the subjects had been maintained and fed in it over a considerable period of time. Additionally, it should be noted that the water volumes in these bowls were considerably larger than those used for rewarding purposes in the study by Best. It is therefore felt that the response

instability is due to some factor other than an inadequate positive reinforcement for maze running.

Further examination of the results suggests that the maze experience in general and not the training treatment is responsible for the changing response preference. First, the subjects of both the trained and the control groups displayed the choice inconsistency. Second, both groups required an equal amount of prodding to run the maze, their average latencies were similar, and the number of turns, withdrawals, and incomplete choices were also similar. A possible explanation may be that the repeated presentation of trials to a subject offsets the positive reinforcement effect of maze-escape. The subjects were placed in a generally negative situation and within that situation performed a response that led to attainment of the home bowl. This positive experience of bowl reinforcement was short-lived, however, because eventually the subjects received further trial treatments. The continued negative reinforcement of repeated maze trials might have induced the animals to attempt another route of escape, bringing about a shift in their turning preference.

In any event, the use of the right-left discrimination with existing training techniques would present some problems in attempts to replicate the regeneration and cannibalism effects. It could never be determined, for instance, if the effect of an injection or other manipulation were the acquisition (or loss) of memory or merely the tendency on the part of the planarian to reverse its response preference. Experiments now in progress are designed to determine (a) the degree of stabilization that can be expected in other types of discriminations and (b) factors that might facilitate response stabilization.

### E. SUMMARY

The ability of planarians to perform a right-left discrimination has been examined. It was observed that 10 of 17 trained subjects were able to achieve a criterion of nine out of 10 correct responses, while none of the 16 control animals succeeded in attaining the same level of performance. However, the subjects of both the trained and the control groups displayed a marked tendency to shift their turning preference, suggesting that some factor peculiar to the maze situation prevents response stabilization.

### REFERENCES

1. BEST, J. B. Protopsychology. *Sci. Amer.*, 1963, **206**, 55-64.
2. BEST, J. B., & RUBENSTEIN, I. Maze learning and associated behavior in planaria. *J. Comp. & Physiol. Psychol.*, 1962, **55**, 560-566.
3. CORNING, W. C., & JOHN, E. R. Effect of ribonuclease on retention of conditioned response in regenerated planarians. *Science*, 1961, **134**, 1363-1364.

4. CORNWELL, P. An attempted replication of studies by Halas et al. and by Thompson and McConnell. *Worm Runner's Digest*, 1961, **3**, 91-98.
5. EDWARDS, A. L. Statistical Methods for Behavioral Sciences. New York: Rinehart, 1960. Pp. 278-282.
6. ERNHART, E. N., & SHERRICK, C. Retention of a maze habit following regeneration in planaria (*D. maculata*). Paper read at Midwestern Psychological Association, May, 1959.
7. GRIFFARD, C. D. Classical conditioning of the planarian *Phagocata gracilis* to water flow. *J. Comp. & Physiol. Psychol.*, 1963, **56**, 597-600.
8. HALAS, E. S., JAMES, R. L., & KNUTSON, C. S. An attempt at classical conditioning in the planarian. *J. Comp. & Physiol. Psychol.*, in press.
9. HALAS, E. S., JAMES, R. L., & STONE, L. A. Types of responses elicited in planaria by light. *J. Comp. & Physiol. Psychol.*, 1961, **54**, 302-305.
10. HUMPHRIES, B. Maze learning in planaria. *Worm Runner's Digest*, 1961, **3**, 114-116.
11. HYMAN, L. H. The Invertebrate: Platyhelminthes and Rhynchocoela. New York: McGraw-Hill, 1951. Pp. 211-214.
12. JACOBSON, A. L. Learning in flatworms and annelids. *Psychol. Bull.*, 1963, **60**, 74-94.
13. LEE, R. N. Conditioning of a free operant response in planaria. *Science*, 1963, **139**, 1048-1049.
14. MCCONNELL, J. V. Memory transfer through cannibalism in planarians. *J. Neuropsychiat.*, 1962, **3**, S42-S48.
15. MCCONNELL, J. V., JACOBSON, A. L., & KIMBLE, D. P. The effects of regeneration upon retention of a conditioned response in the planarian. *J. Comp. & Physiol. Psychol.*, 1959, **52**, 1-5.
16. MCCONNELL, J. V., JACOBSON, R., & MAYNARD, D. M. Apparent retention of a conditioned response following total regeneration in the planarian. Paper read at American Psychological Association, Cincinnati, Ohio, 1959.
17. THOMPSON, R., & MCCONNELL, J. V. Classical conditioning in the planarian, *Dugesia dorotocephala*. *J. Comp. & Physiol. Psychol.*, 1955, **48**, 65-68.
18. WESTERMAN, R. A. Somatic inheritance of habituation responses to light in planarians. *Science*, 1963, **140**, 676-677.
19. ZELMAN, A., KABAT, L., JACOBSON, R., & MCCONNELL, J. V. Transfer of training through injection of "conditioned" RNA into untrained planarians. *Worm Runner's Digest*, 1963, **5**, 14-21.

Department of Biophysics  
Michigan State University  
East Lansing, Michigan



## CHILDHOOD INTELLIGENCE OF FUTURE SCHIZOPHRENICS AND NEIGHBORHOOD PEERS\*<sup>1</sup>

*Department of Psychology, Western Reserve University*

GEORGE W. ALBEE, ELLEN A. LANE, AND JEANETTE M. REUTER

### A. INTRODUCTION

While many investigators have reported inferior intellectual performance by schizophrenics, little is known about their premorbid intelligence level. Winder (3), in his review of the subject, concluded that the evidence for or against the existence of premorbid intellectual deficit in schizophrenia was tenuous, but that the question was of genuine importance.

In a previous paper (1), it was demonstrated that schizophrenics show no decline in intelligence-test performance between early childhood tests and tests administered during their adult psychosis. Although schizophrenics score significantly lower than average while psychotic, they had performed as much below average while school children, many years before the recognized onset of the disorder. This previous study used schizophrenic subjects from publicly supported hospitals, with the further restrictions that these subjects had spent their childhoods in nonsuburban areas of a large city and had attended public schools.

Because schizophrenia is much more prevalent in the lowest socioeconomic groups (2), it would be possible to attribute the observed inferior performance on intelligence tests to the disadvantages known to affect test performance in socially deprived groups. Therefore, conclusions about intellectual deficit in schizophrenia require controls of unaffected groups from the same socioeconomic levels.

The present study reports comparisons between the childhood intelligence-test performance of adult schizophrenics and control groups of subjects who were children in the same classes in the same schools and from the same neighborhoods as the future schizophrenics. If schizophrenics-to-be score below test norms largely because of the disadvantages attributable to socioeconomic status, they should not differ from children in their own school

\* Received in the Editorial Office on March 9, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The research reported in this paper is supported by Grant M-5186 from the National Institute of Mental Health, U.S. Public Health Service, Department of Health, Education, and Welfare.

classes and neighborhoods. If, on the other hand, they score low because of true premorbid intellectual deficit related to a long-term schizophrenic process, they should be noticeably poorer in performance than their same-neighborhood peers.

### B. METHOD

The subjects were 178 adult schizophrenics who were diagnosed at any of several state and VA hospitals and clinics in the Cleveland area and who as children had attended Cleveland Public Schools.<sup>2</sup> Approximately 35 per cent of the subjects were Negro and about 20 per cent were female.

Since 1937, Cleveland school children, with the exception of extremely limited youngsters placed in special classes, have been given group intelligence tests in second, sixth, and eighth grades. Throughout the years, the Kuhlmann-Anderson Intelligence Test has been administered routinely in the second grade and the Cleveland Classification Test has been given in the sixth grade. The Terman-McNemar Test was administered in the eighth grade during all but a few of these years.

The *IQ* for each subject on each of the tests was found in records kept by the Cleveland Board of Education. For most subjects, scores from all three tests were found. A few subjects had attended second grade before the testing program began; some had been absent or had attended other schools at the particular time a test was given. In the end, there were 122 subjects with a Kuhlmann-Anderson score, 154 subjects with a Cleveland Classification score, and 103 subjects with a Terman-McNemar score.

For each test score found, a record was made of the median and of the quartiles of the particular school and class attended by the subject. The mean *IQ* and the standard deviation on each group test were also obtained for a large sample of the entire Cleveland school population in 1943, a year representative of the time when the preschizophrenic sample was tested. There were 106 elementary schools and 22 junior-high schools in the Cleveland Public School system at that time.

If future schizophrenics are not different from other children of their own socioeconomic level in intellectual performance, their test scores should fall equally often into each of the quarters of their own school classes, even when the future schizophrenics come from those most deprived areas of the city which score significantly below the average of all Cleveland school children.

<sup>2</sup> Our research could not have proceeded without the understanding and cooperation of the Cleveland Public School Officials. We are especially indebted to William B. Levenson, Superintendent; George E. Theobald, Assistant Superintendent; Joseph L. Mazur, Chief of the Educational Research Bureau; Clifford L. Graves, Supervisor of Psychological Services; and Dorothea Tobin, Principal Clerk, School Census.

## C. RESULTS

The mean *IQ*s of children who later became schizophrenic adults were lower on all three group tests than the means obtained on these tests by Cleveland school children in a representative year. All of the differences between means reported in Table 1 are significant beyond the one per cent level of confidence.

TABLE 1  
CHILDHOOD *IQ*s OF ADULT SCHIZOPHRENICS COMPARED TO *IQ*s OF ALL CLEVELAND SCHOOL CHILDREN TESTED ABOUT THE SAME TIME

Parameter	Second grade		Sixth grade		Eighth grade	
	Schizophrenic	Normal	Schizophrenic	Normal	Schizophrenic	Normal
<i>N</i>	122	2,613	154	4,166	103	4,960
<i>IQ</i>	93.9	100.8	89.4	99.4	90.6	99.6
<i>SD</i>	11.5	10.1	15.0	13.7	16.2	17.2

As was expected, however, many of the subjects had attended schools located in deprived areas in which children have tested consistently below averages made by pupils in all Cleveland schools. The differences reported in Table 1, then, may represent only expected ones for comparisons between certain neighborhood-school populations and the city-wide population of school children. Therefore, the next comparison, that of future schizophrenics with peers from their own neighborhood, affords a more meaningful and conclusive answer to the question of early deficit.

When each subject's score was placed in the appropriate quarter of his own school class and the results were tested against the hypothesis of equal distribution, a chi-square test yielded differences between the observed and

TABLE 2  
PERCENTAGES OF *IQ* SCORES OF FUTURE SCHIZOPHRENICS WITHIN  
QUARTERS OF THEIR OWN SCHOOL CLASSES\*

School grade	<i>N</i>	Quarters*			
		Fourth	Third	Second	First
Second	122	54.1	19.7	16.4	9.8
Sixth	154	50.0	19.5	13.6	16.9
Eighth	103	41.7	23.3	14.6	20.4

\* The expected frequency for each cell is 25.

the expected frequencies significant at the .001 level of confidence. As is evident from Table 2, a large majority of subjects are below the median of their own classmates and a large proportion of these were in the lowest quarter for all three grades of school.

## D. DISCUSSION

Children who later became schizophrenic adults scored consistently below their own school classmates in second, sixth, and eighth grade. They not only scored lower in *IQ* than children from the whole city, as might be expected of children mainly from deprived areas, but they were also significantly below peers from their same neighborhood schools, even during early childhood. Although it is possible that the socioeconomic level of the preschizophrenic children was not always equal to the median socioeconomic level of others in their own class and school, there is reason to believe there could not have been much difference, especially for elementary schools. Small neighborhood schools in the central areas of this large metropolis during the 30s and early 40s were represented by remarkably homogeneous populations in regard to race, nationality, and occupation. It seems reasonable to conclude, therefore, that intellectual deficit related to a schizophrenic process is present during the childhood of adult schizophrenics, many years before these individuals are recognized and certainly before they are diagnosed as schizophrenic.

## E. SUMMARY

In order to demonstrate that premorbid intellectual deficit in schizophrenia is not merely the result of low socioeconomic backgrounds, intelligence-test scores of 178 children who later became schizophrenic adults were compared with both city-wide averages and with scores of controls from their own school classes and neighborhoods. Future schizophrenics are significantly lower in intellectual level in second, sixth, and eighth grades of school than are their neighborhood school peers.

## REFERENCES

1. ALBEE, G. W., LANE, E. A., CORCORAN, C., & WERNEKE, A. Childhood and intercurrent intellectual performance of adult schizophrenics. *J. Consult. Psychol.*, 1963, 4, 364-366.
2. HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social Class and Mental Illness*. New York: Wiley, 1958.
3. WINDER, C. L. Some psychological studies of schizophrenics. In D. D. Jackson (Ed.), *The Etiology of Schizophrenia*. New York: Basic Books, 1960. Pp. 191-292.

Department of Psychology  
Western Reserve University  
Cleveland, Ohio



## SOME NONPSYCHOLOGICAL VARIABLES IN THERAPY DEFECTION IN A CHILD-GUIDANCE CLINIC\*

*Institute for Juvenile Research, Chicago, Illinois*

---

ROBERT WILLIAMS AND ROBERT H. POLLACK

---

### A. INTRODUCTION

The current literature contains several studies on attrition in outpatient clinics for children and for adults. In general, these studies have relied on psychological traits of the patients as probable determinants of attrition (1, 2, 10, 11). Other studies have focused on sociological traits, but have confined their attention to socioeconomic variables and the relation of class to mental illness (4, 5, 6, 9, 13, 14). Only Levitt's studies (7, 8) investigated nonpsychological parameters of attrition in a child-guidance-clinic population.

The present investigation focused exclusively on some common, readily available, nonpsychological variables of the patient and his family that might affect treatment and defection. The aim was to present empirical data on the subjects rather than to attempt to predict who would defect.

There are certain methodological advantages in the demographic approach to the study of therapy defection. The data observed are unequivocal; they were readily available as a routine part of intake procedures. For this study, case records sufficed to provide all the data, and additional interviews with patients were not necessary.

Before presenting the methodology and procedure used in collecting the data, a brief description of clinic procedures employed at the Institute for Juvenile Research (IJR) is offered by way of clarifying and defining the subgroups used in this study.

The IJR is a tax-supported, free outpatient-clinic for children under the age of 18 years. Patient referrals come from various sources: self-referral, social agencies, schools, physicians and, to a lesser extent, the courts. The clinic procedures consist of a social-history interview with the parents, a psychological examination of the patient, and a psychiatric interview with the patient and his parents. The IJR also maintains an electroencephalograph laboratory in which children with suspected organic damage are examined.

---

\* Received in the Editorial Office on March 11, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.



As a routine part of an application for clinic services at the IJR, parents are asked for consent to contact their family physician, schools, welfare agencies, and other sources for information on the patient's health and adjustment. When all relevant material is gathered, a staff meeting is held to discuss the patient, in order to make a preliminary diagnosis and to recommend disposition of the case. The clinic's findings and recommendations are discussed with the parents in a completion interview. Because of the limited number of families for whom psychotherapy is the ideal disposition and because of the even more limited number of therapists available, only about 20 per cent of the cases seen for diagnosis are accepted for treatment at the IJR. (In the present study, covering a six-year period, 22.5 per cent of the cases seen for diagnosis were accepted for treatment, and 35.2 per cent of these cases remained in treatment until completion of therapy.) Therapy may be offered to one or more family members. While a variety of arrangements may be made, usually a mother and a child are each assigned to a different therapist, each of whom provides weekly interviews. Depending on the current case load, families may have to wait from a few weeks to several months before they can be seen by a staff member.

During the six years covered by this study, certain types of cases were rarely referred to the IJR and were even more rarely offered psychotherapy: the brain-damaged, the mentally retarded, and the psychosomatically ill and psychotic children. All these constitute an insignificant proportion of the sample utilized in this study.

## B. METHOD

The cases selected were seen at the IJR between 1951 and 1957. In about 9 per cent to 12 per cent of the cases used in this study, only the patient's mother received treatment. Just the same, data were available only for the child, and the demographic characteristics reported are exclusively those of the child—regardless of whether he received therapy or not. It should also be noted that the distribution of the number of cases where only the mother received treatment did not significantly differ in the groups accepted for treatment, indicating that this was not a major variable of defection. Data from the social-history records were coded on IBM cards, and the patients were divided into the following groups: (a) those who received treatment until the clinic decided it was no longer necessary (Treatment Group, T); (b) those who defected from treatment once it had begun (Defector Group, D); (c) those who were accepted for treatment and were notified, but who failed to show up for any treatment interviews (Failure Group, F);

(*d*) those who refused treatment after being accepted by the clinic for therapy (Refuser Group, R); and (*e*) those who received a diagnostic examination only and lost contact with the clinic before a treatment decision could be made (Closure Group, C). There was no overlap between groups.

The five subgroups (T, D, F, R, and C) were combined with the remainder of the clinic population seen from 1951 to 1957, and all were examined on the nonpsychological variables to establish population parameters for all cases seen at the IJR. The "remainder" included all cases that could not be included in our original subgroups; i.e., cases that were considered unacceptable for treatment at the IJR, cases that were referred elsewhere, cases for whom treatment was not considered necessary, cases where the patient lacked sufficient assets for therapy, and cases that were considered to have unalterable and serious defects. Cases that failed to complete the diagnostic work-up before dropping out of the clinic were included in this group, also. All these constituted the Population Group, P.

There was no attempt to match the groups on the usual variables of age, sex, race, etc., because these were the variables to be studied.

Because it was the authors' intent to focus on the demographic variables, it might seem reasonable that the psychological variables should have been held constant. They were not, however. Due to the current unreliability of psychiatric diagnoses of neurotic disorders, the limited number of cases presenting complaints recorded, the exclusion of extremely disturbed cases, the amount of diagnostic-staff turnover, and the limited diagnoses recorded (usually mixed neuroses or behavior disorder), it was felt that equating the groups by diagnoses would be unreliable and inaccurate. Therefore, it was assumed that disturbances that existed in the cases were randomly distributed over all the groups and that, psychologically, the groups were equivalent insofar as some perceptible behavior or social problem had originally produced clinic contact.

Tallies for each group were made for the following variables: age at first diagnostic examination; sex of patient; the presence of both natural parents in the home versus the presence of only one natural parent; number of patient's siblings; birth-order of patient (oldest, middle, or youngest); source of referral; type of school attended (public versus parochial); grade level at the time of the first diagnostic examination; both parents' religion; and occupational class of the principal family wage earner.

Because the majority of the variables were discrete, a chi-square analysis was done on all variables, comparing the population parameters with the distribution of the five selected groups and comparing the distributions for

each pair of groups to measure the significance of the differences among and between the distributions of the groups. The variables of age, grade, and number of siblings were divided into discrete groups for the chi-square analysis.

### C. RESULTS

Of the 65  $\chi^2$ s done between the five subgroups and Group P, nine were significant at the .05 level, and 30 at the .01 level of confidence. The probability of obtaining this many significant observations out of 65 comparisons is less than .001 (12).

TABLE 1  
CHI-SQUARE VALUES FOR THE COMPARISON OF SUBGROUPS AND POPULATION GROUPS  
ON 13 NONPSYCHOLOGICAL VARIABLES

Variable	P vs. T	P vs. D	Comparison P vs. F	P vs. R	P vs. C
Age	10.84**	2.15	6.15*	.17	1.03
Sex	.25	6.40*	.06	.69	3.56
Race	8.42**	7.76**	4.29*	6.31*	14.55**
Home	6.28*	11.22**	11.85**	†	13.79**
Siblings					
Number	31.97**	18.34**	8.61*	1.26	4.67
Sex	11.59**	11.00**	8.21*	1.90	4.07
Birth-order	18.89**	16.30**	4.82	2.01	1.32
Referral	22.67**	28.02**	17.74**	9.14*	11.43**
School	.27	.33	.10	2.80	.26
Religion					
Father's	64.34**	25.26**	15.48**	2.01	14.58**
Mother's	62.51**	30.62**	16.98**	2.19	10.03**
School grade	38.70**	33.90**	13.82**	6.83*	1.42
Occupation	24.24**	9.34	7.86	1.73	17.88**

\* Indicates  $\chi^2$  values between the .05 and the .01 levels of confidence.

\*\* Indicates  $\chi^2$  values less than or equal to the .01 level of confidence.

† Indicates that one expected frequency in this comparison was too small to make a valid analysis.

While the value of  $\chi^2$  indicates how far a sample distribution diverges from an expected distribution, it does not tell in which direction this divergence occurs. The direction of divergence is especially important when several groups are compared: e.g., where one group may differ from another in one way and differ from a third group in the opposite direction.

During the comparison of each of the groups separately with P, it was found that Group T had more cases under 7 and over 14 years of age; had fewer Negroes; had fewer cases from homes with only one natural parent

present; had more two-children families and few families with four or more children; had more first- and last-born children in therapy; had more personal and fewer court referrals; had more Jewish and fewer Catholic parents; had fewer children not in school and more in primary school; and had more middleclass occupations than would have been expected from the distribution of these characteristics in P.

Group D was distinguished by displaying a sex difference: there were significantly more females in this group. Also, Negroes were underrepresented; there were fewer cases from homes with one parent present; there were more cases from two-children families and fewer from families with four or more children; there were fewer cases not in school and more in primary school and high school; the patient was most often first-born, while middle- and last-born children were underrepresented; there were more personal referrals and fewer from the courts; and there were more cases with Jewish parents and fewer with Protestant parents than would have been expected by comparison with Group P.

Group F more frequently had children over 8 years of age; had fewer Negroes; had more cases with only the father present in the home and less with only the mother present in the home; had fewer families with four or more children; had fewer children not in school; was most often self-referred; was less often referred by a social agency or the courts; and had more Jewish parents, as contrasted with P.

Group R tended not to differ from Group P, except in regard to race (fewer Negroes); source of referral (more from school, less from personal sources); and grade level (fewer cases not in school). In all other respects, members of Group R were most like the total, unselected clinic population.

Group C had an overrepresentation of Negroes; had more instances of only one parent in the home; had more court referrals; had more Catholic fathers and fewer Jewish parents; was underrepresented by middle-class occupational groups; and was overrepresented by low-income occupational groups.

To summarize, Groups T, D, and F—i.e., three of the four groups selected for treatment—were similar with respect to the direction of case distribution on the variables of race, family size (indicated by the number of siblings), and parental religion. Negroes were underrepresented, families were small, and more Jewish families were present than would have been expected by their known distribution in Group P.

Possible hypotheses about attrition can be made if the groups are compared with the total clinic population. The T Group is least like this population



on 11 of the 13 variables, while the other groups are increasingly more like the general population. This is most interesting when one considers that the general population is predominantly composed of cases rejected by the clinic for psychological reasons.

Of the 130  $\chi^2$ s done between all pairs of groups on the 13 variables, 13 were significant at the .05 level of confidence, and 29 at the .01 level or less. The probability of obtaining this many significant observations out of 130 comparisons by chance alone is less than one out of one thousand (12).

Again, the  $\chi^2$ s do not indicate the direction of divergence between the observed and the expected distributions between groups. As before, a detailed analysis of the results is given in terms of variables rather than groups.

The variable of age was divided into three groups: zero to 7, 8 to 13, and 14 years of age and over. This variable distinguished Group T from Group C. Group T was not particularly over- or underrepresented by any age group when compared with Group C. Group C was underrepresented by the zero-to-7-year-olds and markedly overrepresented in the over-14-year-olds, as compared with Group T.

The sex of the patient failed to distinguish between any groups.

The distribution of Negroes and whites between groups pointed out differences between all the treatment selection groups (T, D, F, and R) and Group C. In all cases, whites were underrepresented and Negroes were overrepresented in cases closing before intake procedures were completed. In general, there was no specific overrepresentation of whites in the other groups, but Negroes were consistently underrepresented in all groups except C.

The presence of both natural parents in the home, as opposed to only one parent, also distinguished Group C cases from all other groups. Group C was consistent in that only one parent in the home (usually the mother) predominated in this group. Also, the frequency of both natural parents present in the home was less than would have been expected by chance. The opposite relation was true for Groups D, F, and R, while Group T had a strong underrepresentation of only the father in the home when compared with Group C.

Group T was distinguished from Group C on the variable of family size. When compared with Group C, Group T came from smaller families.

During the comparison of Group D and Group F with Group C, it was found that Group C also was overrepresented by large family cases, while Groups D and F represented cases with one sibling or with no sibling of the patient.



TABLE 2  
CHI-SQUARE VALUES FOR THE COMPARISON OF ALL SUBGROUPS ON 13 NONPSYCHOLOGICAL VARIABLES

Variable	T vs. R	T vs. C	D vs. R	D vs. C	Comparison				T vs. F	T vs. D	F vs. D
					F vs. R	F vs. C	R vs. C				
Age	1.38	5.99*	.10	2.52	1.75	3.69	.75		6.76*	1.81	2.63
Sex	.85	2.50	3.24	.17	.36	3.40	3.68		.22	2.01	3.01
Race	1.03	28.82**	1.75	25.48**	1.89	17.94**	18.62**		.23	.16	1.34
Home	2.39	19.25**	1.56	16.99**	1.89	12.61**	13.77**		8.16*	5.13	1.34
Siblings <sup>a</sup> Number	4.16	23.61**	3.11	14.71**	.89	10.47*	3.58		3.08	8.85*	4.02
Sex	.86	98.97**	.32	9.72**	.40	8.30*	3.19		.64	.81	.04
Birth-order	9.08*	9.39**	7.31*	8.20*	5.25	4.71	.59		2.19	.35	2.16
Referral	11.40**	6.96	13.25**	8.14*	10.62*	6.98	.81		2.62	1.06	.61
School	2.56	.09	2.59	.09	2.82	.09	2.71		.00	.00	.00
Religion											
Father's	17.84**	39.91**	10.23**	25.45**	8.90*	23.79**	3.45		3.97	4.32	.21
Mother's	17.88**	32.13**	11.79**	24.98**	9.75*	20.49**	1.88		3.10	4.39	.58
School grade	.75	4.64	1.13	5.09	.42	1.08	1.16		2.19	2.61	2.23
Occupation	1.23	29.96**	1.05	22.93**	3.02	24.79**	11.86*		10.56	5.59	6.57

\* Indicates  $\chi^2$  values between the .05 and .01 levels of confidence.

\*\* Indicates  $\chi^2$  values less than or equal to the .01 level of confidence.

The sex of the patient's sibling offered a statistical artifact: the  $\chi^2$  value for Group T *vs.* Group C was extraordinarily high. This was because of the disparity in family size and because of the fact that this variable had three categories (i.e., siblings of the same sex, opposite sex, and of both sexes), so that the probability of having one sibling or no sibling of either sex was infinitesimally small. In comparison with Groups D and F, Group C was overrepresented in all three categories, again reflecting the role of family size in this variable.

The patient's birth-order distinguished Group T from Groups R and C, and Group D from Groups R and C. A slight effect of family size was also indicated here, because Group T was somewhat underrepresented in cases where the youngest child was seen as the patient. Most of the variance was due, however, to the extreme overrepresentation of oldest and youngest children in Groups R and C, each of which showed a strong proportion of older and younger patients. There was no marked tendency for either Groups T or D to favor middle-born children. It should be noted that while the oldest- and the youngest-born children were clearly defined in the samples, the middle-born child was not. In families of more than three children, any child that was not the first- or the last-born was considered middle-born. The ambiguity of the definition of the middle-born child might have something to do with the lack of results in this area.

Informally, the source of clinic referral has been considered a predictor of therapy defection. It would have seemed that voluntary referrals would be more likely to remain in therapy than those referred by schools or by courts. It was found, however, that the source of referral distinguished Groups T, D, and F, members of which were more often self-referred, from Group R, which had more school referrals. The same relationship was true in the comparison of Group D with Group C.

The type of school attended by patients did not distinguish among the groups.

An examination of Table 2 reveals an interesting relationship between the groups compared. Group T *vs.* Group D, T *vs.* F, and D *vs.* F were not statistically different from one another, as measured by the demographic variables. When only the significant chi-square values for these groups are considered, the number reaching the .05 level of significance can be considered a chance occurrence only (12). The greatest number of significant differences occurs between these groups and Group C. In the light of non-psychological variables, these results indicated that Group C was least

similar to the groups accepted for treatment and to those cases in which treatment was not offered at the IJR.

Parental religion discriminated six of the 10 pairs of groups (see Table 2). Because there was no substantial difference between the variables of the mother's and the father's religion, and because there were extremely few cases in which the parents had different religions, these two variables (mother's and father's religion) are treated as one variable in this discussion. Generally, Jewish families were overrepresented in Groups T, D, and F, as compared to Groups R and C. The latter were underrepresented by Jewish families and overrepresented by Catholic families when compared with Groups T, D, and F. This underrepresentation of Jewish families in the nontreatment groups was consistently observed. A disproportionate number of Catholic families was observed in the Group C comparisons only; i.e., overrepresentation of Catholic families occurs in Group C only.

Although academic level was not a discriminating variable, socioeconomic class, as indicated by the occupation of the principal wage earner, distinguished all other groups from Group C. In all comparisons, Group C had a larger proportion of low-income occupation cases and was underrepresented by cases holding middle-income jobs.

#### D. DISCUSSION

During the comparison of five clinical groups on 13 nonpsychological variables, it was revealed that there were no unique variables that distinguished one group from another in terms of therapy defection. Parental religion, however, successfully differentiated the groups accepted by the clinic for therapy (T, D, and F groups) from those groups that directly or indirectly refused clinic aid (R and C groups). The only other variables that differentiated nearly as many groups were the home situation and the number of the patient's siblings. None of the variables distinguished Group T from Group D or from Group F. A distinction between these three groups would obviously have been most important in the practical discrimination of a clinic, because an equal amount of intake time is spent on these groups. Family size, as reflected by the number of the patient's siblings, did differentiate T cases from D cases, and the presence of both natural parents in the home identified T cases as opposed to F cases. This would seem to indicate that T cases generally came from small, stable families. However, when the other groups were compared, the same conclusion could be made about D cases and F cases, as opposed to early C cases.

In conclusion, it may be said that the variables used here are capable of differentiating various levels of closing (but not in a systematic or predictive way). This analysis has shown that cases accepted for treatment bear a strong similarity to one another and bear less similarity to those who refuse therapy or lose clinic contact at an early stage.

While this study has focused on groups and variables within these groups, the fact was not ignored that the selection of treatment cases is much more a function of clinic policy than of individual choice. It was not possible, however, to measure precisely the effect of clinic selection policies on these different groups. Inferences about these policies might be made from the data (preference for preadolescent, middle-income, white, Jewish, male treatment patients from intact, small families), but considering the statistics used, such inferences are not fully warranted.

The results of the present investigation tend to support as well as to contradict the findings of other investigators. That the middle-income groups are more likely to enter into and maintain clinic contact throughout intake (14) was confirmed. It was not found, however, that socioeconomic class played a significant role in determining whether or not the patient remained in therapy (4, 13, 14). The discrepancy between the current findings and past studies may be due to differences in definitions of class and defection, as well as to differences in samples. The correlation between socioeconomic class, religion, and request for therapeutic aid found in a previous study (9) was also substantiated, with a slight difference. While Roach *et al.* found both Protestant and Jewish families to be overrepresented in his sample, the current investigators found that only Jewish families predominated in the treatment groups (T, D, and F). The apparent difference between the two studies is probably due to the relative size of the nonwhite portions of the respective samples: one per cent in the study of Roach *et al.* (9), and 12 per cent in the present study.

In view of the demographic similarity of all the cases selected for treatment, it is extremely interesting that one study, also done at the IJR (8), failed to find any significant differences in posttherapy adjustment between those who received therapy and those who defected from it.

Hopefully, this study points up the fact that there are more than psychological variables present in the selection of treatment cases in a child-guidance clinic, on the part of either the patient or the clinic, or both. It is also hoped that the role of nonpsychological variables in differentiating the various clinic subgroups has been spotlighted.

## E. SUMMARY

The relationship of nonpsychological or demographic variables to therapy defection in a child-guidance clinic was statistically investigated. Certain differences were found to exist among the treatment-, the failure-, the defector-, and the closure-group, but none proved to be a significant indicator of therapy defection. It was observed, however, that there is an apparent selection of cases accepted for, or accepting treatment, on the basis of demographic characteristics.

## REFERENCES

1. GIBBY, R. G., STOTSKY, B. A., HILER, E. W., & MILLER, D. R. Validation of Rorschach criteria for predicting duration of therapy. *J. Consult. Psychol.*, 1954, **18**, 185-191.
2. GIBBY, R. G., STOTSKY, B. A., MILLER, D. R., & HILER, E. W. Prediction of duration of therapy from the Rorschach Test. *J. Consult. Psychol.*, 1953, **17**, 348-354.
3. HILER, E. W. Initial complaints as predictors of continuation in psychotherapy. *J. Clin. Psychol.*, 1959, **15**, 344-345.
4. HOLLINGSHEAD, A. B., & REDLICH, F. C. Social Class and Mental Illness: A Community Study. New York: Wiley, 1958.
5. IMBER, S. D., NASH, E. H., JR., & STONE, A. R. Social class and duration of psychotherapy. *J. Clin. Psychol.*, 1955, **11**, 281-284.
6. INMAN, A. Attrition of child guidance: A telephone follow-up study. *Smith Coll. Stud. Soc. Wk.*, 1956, **27**, 34-75.
7. LEVITT, E. E. A comparison of "remainders" and "defectors" among child clinic patients. *J. Consult. Psychol.*, 1957, **21**, 316.
8. ———. A follow-up study of cases treated at the Illinois Institute for Juvenile Research. Unpublished manuscript, Chicago, Illinois, 1957.
9. ROACH, J. L., GURRLIN, O., & HUNT, R. G. Some social-psychological characteristics of a child guidance clinic caseload. *J. Consult. Psychol.*, 1958, **22**, 183-186.
10. ROSS, A. O., & LACEY, H. M. Characteristics of terminators and remainders in child guidance treatment. *J. Consult. Psychol.*, 1961, **25**, 420-424.
11. RUBENSTEIN, E. O., & LORR, M. A comparison of terminators and remainders in outpatient psychotherapy. *J. Clin. Psychol.*, 1956, **12**, 345-349.
12. SAKODA, J. M., COHEN, B. H., & BEALL, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, **51**, 172-175.
13. STEVENS, S. An ecological study of child guidance intake. *Smith Coll. Stud. Soc. Wk.*, 1954, **25**, 73-84.
14. TUCKMAN, J., & LAVELL, M. Social status and clinic contact. *J. Clin. Psychol.*, 1959, **15**, 345-348.

*Institute for Juvenile Research*  
608 South Dearborn Street  
Chicago, Illinois 60605



## CEREBRAL DOMINANCE AND AUDITORY ASYMMETRY\*<sup>1</sup>

*Worcester State Hospital; and Institute of Human Development, Clark University*

---

ROBERT D. PALMER

---

### A. INTRODUCTION

The present study is concerned with the effect of cerebral dominance upon the relative efficiency of the two ears. Several aspects of this problem require examination. Though the central auditory pathway involves both ipsilateral (uncrossed) and contralateral (crossed) connections from ear to cortex, evidence from several sources suggests that the contralateral pathways are more efficient than the ipsilateral. Physiological animal research by Tunturi (29) and by Rosenzweig (24), for example, clearly indicates that the contralateral auditory pathway is stronger in terms of amplitude of evoked cortical response.

The greater efficiency of the contralateral pathways has also been demonstrated in cases of unilateral temporal lobe dysfunction in man. By way of clarification, note that if the performance of one temporal lobe is seriously impaired, the ear on the opposite side will be left only with its ipsilateral connection, whereas the ear on the same side will retain only its contralateral connection. If the contralateral pathways are in fact stronger or more efficient than the ipsilateral pathways, performance of the ear contralateral to the malfunctioning temporal lobe should show the greatest impairment. Standard audiometric techniques reveal but a slight loss for high frequencies in the ear contralateral to unilateral temporal lobectomy (25), or no loss at all (12). Contralateral auditory deficit has been clearly demonstrated, however, under special experimental conditions. Bocca, Calearo, Cassinari, and Migliavacca (1), for example, report that recognition of accelerated speech, or of words distorted by a low-pass filter, was significantly impaired on the side contralateral to a temporal-lobe tumor, though normal speech was not so affected. Sinha (25) found contralateral deficit after temporal lobectomy for seizures when words were presented with white noise, but not when words were pre-

---

\* Received in the Editorial Office on March 12, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The writer is indebted to John A. MacGahan and Jay Melrose of the Boston VA Outpatient Clinic for invaluable assistance in the instrumentation of this research. He wishes also to thank Donald M. Broverman and William Vogel for many helpful comments during the preparation of the paper. This work was supported in part by U.S. Public Health Service Grant M-5773.

sented alone. Jerger and Mier (13) observed auditory deficit contralateral to a lesion of the auditory cortex under conditions in which an irrelevant conversation was presented to the ipsilateral ear. More recently, Kimura (15), using a technique introduced by Broadbent (2), presented different (spoken) digits to the two ears either simultaneously or in rapid succession. Impaired recognition for digits arriving at the ear contralateral to unilateral temporal lobectomy was found only under the condition of simultaneous presentation, in which some form of rivalry or competition between the two pathways was presumably involved.

Kimura's study revealed, in addition, that for all groups of patients studied, regardless of site of lesion, the right ear was more efficient than the left, prior to operation. Similar results have been obtained with normal adults, both by Kimura (personal communication) and by Bryden (3). Kimura (17) has also obtained this "right-ear effect" with children at different age levels, including children as young as age 4. Because of the apparently greater efficiency of the contralateral auditory pathways, Kimura attributed the "right-ear effect" to the fact that the left hemisphere is dominant for speech representation in most individuals. Kimura (16) has confirmed this interpretation by demonstrating greater left-ear efficiency for subjects known to have speech representation in the right hemisphere and greater right-ear efficiency for subjects having speech representation on the left. That is, stimuli arriving at the ear contralateral to the dominant hemisphere for speech were more efficiently recognized.

Kimura's "right-ear effect" (in left-hemisphere-dominant subjects) would appear to obtain only for verbal stimuli, however. In a preliminary study reported by Milner (20), Kimura tested a group of normals on a nonverbal auditory task involving the presentation of different numbers of clicks (rather than spoken digits) to the two ears simultaneously, with instructions to report the number of clicks arriving at each ear. On this task, a tendency (not significant) was found for the left ear to be more efficient, rather than the right. These findings are consistent with evidence presented by Milner (19, 20), Lansdell (18), and others suggesting the possible superiority of the right hemisphere (or at least the absence of marked left-hemisphere superiority) for certain nonverbal functions.

Studies using techniques of *monaural* stimulus presentation (with normal subjects) have not so far produced consistent, conclusive evidence for a hemisphere effect in audition. Glorig, Wheeler, Quiggle, Grings, and Summerfield (8) obtained lower right-ear thresholds both for speech and for pure-

tone reception, particularly among males and at higher pure-tone frequencies.<sup>2</sup> Females obtained lower right-ear thresholds for high frequencies and lower left-ear thresholds for low frequencies. These differences were relatively slight, but because of the size of the sample (3,465 subjects) the differences tended to be clearly significant. A marked tendency to right-ear superiority is seen, too, in Corso's pure-tone results for older men (7), but is not present in his results for women or younger men. Corso (6) reports slightly greater *left-ear* superiority for speech reception in a mixed group of 73 males and 66 females (significance level not reported). Other investigators have found little or no difference in right- and left-ear thresholds for normal speech (14) or for interrupted or distorted speech (4). Though the Glorig *et al.* finding of greater right-ear efficiency for speech reception is highly significant, statistically, the overall picture clearly is confusing. While monaural presentation may simply not be sensitive to dominance effects, lack of agreement among these latter four studies might conceivably be attributable, in some degree, to a failure to counterbalance carefully for presentation of stimuli to right and left ears—especially since the investigators (Calearo and Antonelli excepted) were concerned only incidentally with right-left threshold differences.

Glorig *et al.* wished primarily to establish population norms for normal hearing, and so did not attempt to measure degree of asymmetry in hearing function as manifested by the individual. This latter kind of measure would seem particularly to be desired when exploring hypotheses related to cerebral dominance, because simple categorization of an individual as either "left-brained" or "right-brained" is considered by many present-day investigators to constitute an oversimplification of the problem of cerebral dominance (22). According to Zangwill (30), for example, "cerebral dominance is in all probability . . . a graded characteristic, varying in scope and completeness from individual to individual." Similar positions have been espoused by Conrad (5), Subirana (28), Goodglass and Quadfasel (9), Hécaen and Piercy (10), and others. If greater right-ear efficiency is to be attributed to the dominance of the left hemisphere for language functions, the degree to which these functions are lateralized in any given individual should bear some correspondence to the degree of threshold discrepancy manifested. One

<sup>2</sup> The high frequencies would seem to be more susceptible than the lows to stress in general. It will be recalled that the slight deficit reported by Sinha (for the ear contralateral to a lesion of the temporal lobe) was for high frequencies. Hearing losses associated with age are greater for high than for low frequencies. So too is the loss that results from noise exposure. It is also reported (8) that among men who had often shot a gun, the left ear showed more loss than the right, particularly at high frequencies.

is on surer ground theoretically, of course, in predicting right-ear superiority for verbal stimuli than for nonverbal stimuli.

The objectives of the present study were (a) to attempt to demonstrate greater right-ear efficiency (i.e., an auditory effect attributable to the language dominance of the left hemisphere) for monaurally presented speech stimuli, and (b) to explore the use of scores reflecting the degree of discrepancy between individuals' thresholds for right and left ears, as such scores achieve theoretical significance as indicators of degree of asymmetry or "dominance" in auditory function.

## B. METHOD

### 1. Subjects

Ss were seen in conjunction with a larger project concerned with lateralized motor development (21). Ss were 55 paid male Harvard undergraduates, ages 18 to 22. They were tested individually. Respondents were not accepted as subjects if they perceived their hearing as other than relatively normal or if they reported ever having had a "running ear," punctured eardrum, serious infection of the inner ear, or specialized medical attention for an ear ailment. None was a war veteran.

### 2. Materials and Procedure

*a. Auditory stimuli.* The test stimuli used in the present experiment were taken from Central Institute for the Deaf Auditory Test W-2 recorded spondee word lists (11). This is a "descending-level" test comprised of six lists of words, each list being a different scrambling of the same 36 spondaic words. All are familiar words ("baseball," "railroad," etc.) and are recorded in such a way as to be equally intelligible at any given intensity level. Only four of the six lists were used in the present experiment. Each list is recorded with "step-down" attenuation. The first three words in each list are recorded at the level of the recorded 1000-cps calibration tone, the next three words are down three decibels, the following three are down six decibels, etc. Test words are presented at intervals of approximately six and one-half seconds and are preceded by the phrase, "Say the word . . ." at an intensity level six decibels above that of the test word itself.

*b. Equipment.* For the present experiment, four of the six lists were transferred from long-playing record to audio tape. Sections of the recorded calibration tone were spliced into the tape at the beginning of each list. A Wollensak tape recorder and a rewired Koss Stereophone Headset (Model SP-3) were used to present the test stimuli to the subject. Calibration was



achieved through use of a VU meter. Use of a rotary selector switch with "shorting" (low-noise) contacts permitted monaural presentation of stimuli to the right or to the left ear (never to both ears simultaneously).

*c. Procedure.* *S* was told that he would hear a series of familiar, two-syllable words, that words would be presented to right and left ears alternately, and that the words would become softer and softer as the test progressed. He was strongly encouraged to guess when in doubt or if he understood only one of the syllables. The recorded calibration tone was checked against the reference level of the meter prior to each of the four descents. To control for practice and fatigue effects, each successive stimulus word was presented to the ear opposite that which had received the previous word (R-L-R-L, etc.). To control for possible inequality in the output of the right and the left headphones, the headset was reversed for two of the four descents.

Threshold scores were calculated for each ear separately. Half credit was given for one syllable correct and (to accommodate for minor distortion introduced in the transfer of spondees from record to audio tape) for several responses shown—on the basis of a frequency tally of inaccurate responses—to be frequently elicited. A discrepancy score was also obtained, based on the absolute difference between *S*'s threshold scores for right and left ears.

### C. RESULTS

For the group as a whole, the mean left-ear threshold is slightly higher than the mean right-ear threshold, in keeping with theoretical expectations and with the Glorig *et al.* findings; the difference is minimal, however, and is not significant as evaluated by the *t* test for correlated means ( $t = 1.296$ ,  $df = 54$ ). No difference was found in the proportion of subjects manifesting right-ear *vs.* left-ear superiority: 26 subjects obtained a higher error (threshold) score for the left ear than for the right, while 27 subjects made more errors on the right ear. Two subjects showed no difference. The present findings clearly offer little support, then, for the expected auditory effect.

Significant relationships were found, however, when the degree of asymmetry in auditory function (or the degree of "deficit" from an individual base line) was taken into account, through consideration of the degree of discrepancy between an individual's thresholds for right and left ears. Those 26 subjects who manifested the expected pattern (poorer acuity for the left ear) were found to have significantly higher discrepancy scores than the 27 subjects with poorer acuity on the right ( $t = 2.313$ ,  $p < .05$ ,  $df = 51$ ). The "left-ear-worse" group averaged 10.3 and 14.8 errors for the right ear and



the left ear respectively, for a mean discrepancy of 4.5. Corresponding figures for "right-ear-worse" subjects are 13.7 and 11.2 errors, for a mean discrepancy of 2.5.

Another way of viewing these data is to observe the correlation of discrepancy scores with right-ear scores and left-ear scores respectively. In the absence of functional differences between the two hemispheres, one would expect the correlation between discrepancy scores and right-ear thresholds to be approximately equal in magnitude to the correlation between discrepancy scores and left-ear thresholds. These two correlations should also be low in magnitude. In the present sample, however, discrepancy scores are found to be fairly strongly associated ( $r = .48$ ) with thresholds for the left ear (the ear contralateral to the nondominant hemisphere), high discrepancies being associated with raised thresholds. Discrepancy scores are but nominally associated ( $r = .06$ ) with thresholds for the right ear, however. The former correlation is significantly different from zero ( $p < .001$ ). Using the formula for nonindependent correlations, the difference between the two correlations themselves is found to be significant ( $t = 3.642$ ,  $p < .001$ ). Though the left-ear threshold-scores have slightly greater variance, the difference between right- and left-ear variances does not begin to approach significance as evaluated by the formula for nonindependent variances ( $t = 0.242$ ). Hence, the present pattern of results cannot be attributed to differences in variances for right and left ears.

#### D. DISCUSSION

The minimal difference found between right-ear and left-ear thresholds in the present study, and the fact that the proportions of individuals manifesting right-ear superiority and left-ear superiority are almost identical, lends further support to Kimura's position that distinctly greater right-ear efficiency is to be expected only under "competitive" conditions in which different stimuli are presented to the two ears simultaneously. We note, however, that those individuals who manifested the expected pattern of poorer efficiency for the left ear (the ear contralateral to the nondominant hemisphere) gave evidence of greater asymmetry in auditory function (or greater "deficit" relative to an individual base line), as indicated by a high discrepancy in thresholds. Expressed differently, unilateral "deficit" was found to be greater when it occurred for the left ear (contralateral to the nondominant hemisphere) than when it occurred for the right ear. It would appear, therefore, that an auditory effect attributable to the dominance of the left hemisphere can in fact be demonstrated with monaural audiometric techniques. The effect ap-

pears, however, to be of a different order from that observed by Kimura, under conditions of simultaneous stimulus presentation, and cannot be as simply explained.

Since the expected proportion of "left-brained" individuals in an unselected population is something less than 100 per cent, knowledge of subjects' dominant hemisphere should lead to (slightly) increased predictability regarding right-ear *vs.* left-ear efficiency, as in the Kimura study (16). So too should knowledge of the degree to which subjects are lateralized for language. That is, if greater right-ear efficiency is in fact to be attributed to the dominance of the left hemisphere for language functions, the degree to which language functions are lateralized in any given individual should bear some correspondence to the degree of threshold discrepancy manifested. The results of the present study provide tentative support for this expectation. It is well established that the vast majority of persons are left-hemisphere dominant for language. Evidence cited above, however, suggests a tendency for the contralateral auditory pathways to be more efficient. It is possible, therefore, that left-ear-dominant subjects in the present study, though not actually right-hemisphere-dominant for language, were in their language functions less completely lateralized to the left hemisphere. To recapitulate, those individuals manifesting right-ear superiority, and in whom language functions were conceivably more strongly lateralized, manifested greater asymmetry in auditory function. This interpretation of our results must be considered tentative, and further research utilizing more direct measures of degree of lateralization or focalization of language functions clearly is desirable.

Interpretation of discrepancy scores is confused by the fact that such scores can be viewed either (*a*) as measures of asymmetry in auditory function (attributable to asymmetries in brain function or to "cerebral dominance"), or (*b*) as measures of hearing loss ("unilateral deficit"). The clinical audiologist, confronted again and again with instances in which disease or injury has resulted in impaired hearing for one or both ears, would of course be more inclined to the latter frame of reference. And in general it does seem reasonable to attribute a high threshold discrepancy in large degree to impaired hearing for one ear (rather, say, than to "enhanced" auditory efficiency for the opposite ear). We note, for example, that in the present study the 25 subjects with the lowest threshold discrepancies obtained mean errors of 10.7 and 11.7 for "best ear" and "worst ear" respectively, while 24 high-discrepancy subjects obtained mean errors of 10.8 and 16.8 for "best" and "worst" ears; i.e., the high-discrepancy subjects clearly have poorer hearing overall. Absolute discrepancy was found to be strongly associated with hearing for "worst ear"

( $r = .59$ ), high discrepancies being associated with poor hearing. It was but slightly associated with hearing for "best ear" ( $r = -.07$ ). Subjects' "worst-ear" scores also showed significantly greater variability than "best-ear" scores.

Logically speaking, however, a high absolute difference in thresholds for the two ears could be due to auditory deficit for one ear or to enhanced acuity for the other. If individuals who are most strongly lateralized for language manifest a greater discrepancy in thresholds, one would hesitate to use the label "unilateral deficit" in describing their hearing. That both "deficit" and "increment" may be involved is suggested by the fact that the 26 "left-ear-worse" subjects in the present study averaged 10.3 and 14.8 errors for the right ear and the left ear respectively, whereas the 27 "right-ear-worse" subjects averaged 13.7 and 11.2 errors. That is, subjects who manifested the expected pattern of poorer efficiency for the left ear, and in whom language functions may conceivably be more strongly lateralized, appear to have achieved a higher threshold discrepancy both by virtue of elevated thresholds on the left and by virtue of lowered thresholds on the right.

It is not clear why greater auditory efficiency contralateral to the dominant hemisphere, or deficit contralateral to unilateral temporal-lobe disturbance, *cannot* be demonstrated with standard monaural audiometric techniques, while differences *can* be demonstrated under experimental conditions involving the distortion or masking of verbal stimuli or involving the presentation of verbal stimuli to the two ears simultaneously. Contralateral deficit conceivably may be elicited only when the auditory pathways are "overloaded" in some sense, or when close attention is required, or when the subject must separate relevant from irrelevant stimulation (a condition that may require auditory integration on a higher cortical level). Also, all such conditions make veridical report more difficult and, by enlarging the threshold region, may conceivably permit more sensitive measurement.<sup>3</sup>

The sensitivity of monaural audiometric measures to dominance effects conceivably might be increased through slight modifications in technique. The W-1 and W-2 recorded spondee word lists in particular—though they perform a useful function in clinical settings as rather global measures of speech reception—possess characteristics that limit their utility as research instruments.

<sup>3</sup> The situation may be analogous to that which appeared to obtain in a visual recognition experiment by Postman and Schneider (23). In that experiment it seemed that a subject's selective sensitivity to different value areas had a poorer chance to show itself (in altered tachistoscopic recognition thresholds) for familiar words, where recognition is fast and the threshold region small, than for relatively unfamiliar words, where recognition builds up over a longer period of time.

Because each list represents a different scrambling of the same 36 words, the subject develops an increasing familiarity with the test stimuli if more than one descent is used in obtaining a threshold. The influence of word familiarity (as established by the Thorndike-Lorge word counts) on tachistoscopic recognition thresholds is, of course, well known. Such relationships have also been demonstrated in studies in which frequency of usage has been more precisely controlled, as by having the subject read and pronounce foreign words or nonsense words a given number of times prior to tachistoscopic presentation of the same words (26, 27). Even were threshold calculations based on the same number of descents for each subject, interindividual threshold differences could be attributable to differential learning or retention of test stimuli during the assessment period, to differences in prior exposure to these words, to interindividual differences in the number of spondee presentations required to establish a threshold (because of the effect on learning and retention of test stimuli), to differences in general verbal facility or vocabulary level, to differences in willingness to report, or to differences in attention, rather than to "auditory acuity" narrowly conceived. Use of a multiple-choice auditory test, by overcoming certain of these obstacles, might permit a more precise statement of the effects of hemisphere dominance on auditory efficiency. The utility of threshold-discrepancy scores in the present study may be attributable in some degree to the fact that these scores utilize the individual as his own baseline, and so are freed to a large extent from the influence of such factors.

#### E. SUMMARY

The present study utilized intraindividual threshold comparisons in assessing the effect of cerebral dominance upon the relative efficiency of the two ears. Only mild support was found for the hypothesis that speech stimuli presented monaurally to the right ear (contralateral to the dominant hemisphere) are more readily recognized than stimuli presented to the left ear. This result lends further support to Kimura's position that distinctly greater right-ear efficiency is to be expected only under "competitive" conditions of simultaneous stimulus presentation. However, though the proportions of individuals manifesting right-ear *vs.* left-ear superiority were almost identical, those individuals manifesting the expected pattern of right-ear superiority (and in whom language functions were conceivably more strongly lateralized) manifested significantly greater asymmetry in auditory function. If greater right-ear efficiency is to be attributed to the dominance of the left hemisphere for language functions, it would of course be reasonable to expect some correspondence between the degree to which these functions are lateralized and the



degree of threshold asymmetry manifested. Further research utilizing more direct measures of degree of lateralization of language functions is needed.

The techniques of intraindividual threshold-comparison utilized in the present study constitute a sensible way of bringing to light an important but elusive effect. They deserve consideration by other investigators concerned with hemisphere effects in audition.

#### REFERENCES

1. BOCCA, E., CALEARO, C., CASSINARI, V., & MIGLIAVACCA, F. Testing "cortical" hearing in temporal lobe tumors. *Acta Otolaryng.* (Stockholm), 1955, **45**, 289-304.
2. BROADBENT, D. E. The role of auditory localization in attention and memory span. *J. Exper. Psychol.*, 1954, **47**, 191-196.
3. BRYDEN, M. P. Ear preference in auditory perception. *J. Exper. Psychol.*, 1963, **65**, 103-105.
4. CALEARO, C., & ANTONELLI, A. R. "Cortical" hearing tests and cerebral dominance. *Acta Otolaryng.* (Stockholm), 1963, **56**, 17-26.
5. CONRAD, K. Über aphasische Sprachstörungen bei hirnerkrankten Linkshändern. *Nervenärzt*, 1949, **20**, 148-154.
6. CORSO, J. F. Confirmation of the normal threshold for speech on C.I.D. auditory test W-2. *J. Acoust. Soc. Amer.*, 1957, **29**, 368-370.
7. ———. Age and sex differences in pure-tone thresholds. *Arch. Otolaryng.* (Chicago), 1963, **77**, 385-405.
8. GLORIG, A., WHEELER, D., QUIGGLE, R., GRINGS, W., & SUMMERFIELD, A. 1954 Wisconsin State Fair Hearing Survey. Rochester, Minn.: American Academy of Ophthalmology & Otolaryngology, 1957.
9. GOODGLASS, H., & QUADFASER, F. A. Language laterality in left-handed aphasics. *Brain*, 1954, **77**, 521-548.
10. HÉCAEN, H., & PIERCY, M. Paroxysmal dysphasia and the problem of cerebral dominance. *J. Neurol. Neurosurg. & Psychiat.*, 1956, **19**, 194-201.
11. HIRSH, I. J., DAVIS, H., SILVERMAN, S. R., REYNOLDS, E. G., ELDELT, E., & BENSON, R. W. Development of materials for speech audiometry. *J. Speech & Hearing Dis.*, 1952, **17**, 321-337.
12. JERGER, J. F. Observations on auditory behavior in lesions of the central auditory pathway. *Arch. Otolaryng.* (Chicago), 1960, **71**, 797-806.
13. JERGER, J. F., & MIER, M. The effect of brain stem lesions on auditory responses of humans. Paper read at First Annual Meeting of Psychonomic Society, Chicago, 1960.
14. JERGER, J. F., CARHART, R., TILLMAN, T. W., & PETERSON, J. L. Some relations between normal hearing for pure tones and for speech. *J. Speech & Hearing Res.*, 1959, **2**, 126-140.
15. KIMURA, D. Some effects of temporal-lobe damage on auditory perception. *Can. J. Psychol.*, 1961, **15**, 156-165.
16. ———. Cerebral dominance and the perception of verbal stimuli. *Can. J. Psychol.*, 1961, **15**, 166-171.
17. ———. Speech lateralization in young children as determined by an auditory test. *J. Comp. & Physiol. Psychol.*, 1963, **56**, 899-902.
18. LANDSDELL, H. Laterality of verbal intelligence in the brain. *Science*, 1962, **135**, 922-923.



19. MILNER, B. Psychological defects produced by temporal lobe excision. *Publ. Assoc. Res. Nerv. Ment. Dis.*, 1958, **36**, 244-257.
20. ———. Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric Relations and Cerebral Dominance*. Baltimore: Johns Hopkins Press, 1962.
21. PALMER, R. D. Hand differentiation and psychological functioning. *J. Personal.*, 1963, **31**, 445-461.
22. ———. Development of a differentiated handedness. *Psychol. Bull.*, in press.
23. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271-284.
24. ROSENZWEIG, M. R. Representations of the two ears at the auditory cortex. *Amer. J. Physiol.*, 1951, **167**, 147-158.
25. SINHA, S. P. The role of the temporal lobe in hearing. Unpublished Master's thesis, McGill University, Montreal, Quebec, 1959.
26. SOLOMON, R. L. Visual duration thresholds as a function of experimentally controlled verbal frequency. Paper read at Eastern Psychological Association, 1951.
27. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition threshold for words. *J. Exper. Psychol.*, 1952, **43**, 195-202.
28. SUBIRANA, A. La Droiterie. *Schweiz. Arch. Neurol. Psychiat.*, 1952, **69**, 321-359.
29. TUNTURI, A. R. A study on the pathway from the medial geniculate body to the acoustic cortex in the dog. *Amer. J. Physiol.*, 1946, **147**, 311-319.
30. ZANGWILL, O. L. Cerebral Dominance and Its Relation to Psychological Function. London: Oliver & Boyd, 1960.

Department of Psychology  
Worcester State Hospital  
Worcester 1, Massachusetts

## PERSONALITY VARIABLES RELATED TO SELF-REINFORCEMENT\*

*Department of Psychology, University of Wisconsin*

---

ALBERT R. MARSTON

---

### A. INTRODUCTION

Several recent studies (4, 5, 8) have examined *self-reinforcement* (sr) as a process relevant to behavioral change. Skinner (11) mentions self-administration of reinforcement without environmental restrictions as one of the ways in which the human organism institutes self-control or shapes his own behavior. Each individual has available a pool of reinforcing stimuli—either physical objects or social reinforcers in the form of verbalizations—but he takes a reinforcer only when the behavior to be reinforced meets his own criteria for reinforcement.

The examination of the self-reinforcement process has involved three areas of consideration: (a) variables influencing the occurrence of self-reinforcement; (b) the reinforcing or fixating effect of self-reinforcement; and (c) the direct, external reinforcement of sr in its own right. The present study is concerned with the first of these areas.

Self-reinforcement can be conceptualized as behavior that, in addition to various situational determinants, has a large component of variance attributable to stable, individual characteristics, or personality variables. Individuals maintain consistent criteria for reinforcing themselves (7). The taking of a positive reinforcement when a criterion has *not* been met, or failing to take a negative reinforcement when a criterion has been met, appears to result in responses (overt and covert) that are associated with aversive stimuli—i.e., with guilt.

The characteristic individual sr pattern may be important in the development of certain types of psychopathology and in aspects of psychotherapy. Extremely low criteria or the absence of criteria for positive sr characterizes the psychopath. In large part, low criteria for negative sr describes the depressed individual. Often an essential phase in successful therapy involves maintenance of behavior change by self-reinforcement after a period of dependence on the therapist for crucial reinforcements.

---

\* Received in the Editorial Office on March 23, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

The purpose of the two experiments described in this report was to examine the relationship of positive *sr* to the personality variables measured by Bass's Orientation Inventory (1), and Rotter's Internal-External Control of Reinforcements Scale (10).

## B. METHOD

### 1. Design

For investigating the Orientation Inventory (ORI), a  $2 \times 3$  factorial design was utilized: sex of subject and three categories from the ORI (self-orientation, interaction-orientation, task-orientation). For Rotter's I-E Scale, a  $2 \times 2$  design was used: sex of subject and high *vs.* low scores on the I-E scale.

### 2. Tests

The ORI is based on Bass's study of group processes (1). It enables one to classify individuals as self-oriented, interaction-oriented, or task-oriented. To each of 27 items, the subject (*S*) chooses the endings he least prefers and the endings he most prefers. For example, one item states, "I like: (a) Personal praise, (b) Cooperative effort, (c) Wisdom."

The self-oriented individual is defined as one whose scores place him in the top quarter of scores based on preference for self-oriented endings *and* in the lower three-quarters of the distribution on each of the other two variables. Such a person is overly concerned with rewards for himself at the cost of smooth interpersonal relationships and task accomplishments.

The interaction-oriented individual is defined as one whose scores place him in the top quarter of scores based on preference for interaction-oriented endings *and* in the lower three-quarters of the distribution on each of the other two variables. Such a person sacrifices self-aggrandizement and task completion for harmonious interactions.

The task-oriented individual is defined as one whose scores place him in the top quarter of scores based on preference for task-oriented endings *and* in the lower three-quarters of the distribution on each of the other two variables. Such a person is most concerned with "getting the job done" and less concerned with rewards and interpersonal harmony.

Among other variables, those found related to scores on the ORI are styles of group behavior, type of profession (3), and the seeking of social approval in dyadic interactions (6).

The I-E test is based on Rotter's theories concerning expectation of reinforcements (9). It measures the degree to which an individual sees his rein-

forcements as due to an uncontrollable, external world rather than as due to himself as the controller of his reinforcements. A typical item on the 29-item scale is as follows: "I more strongly believe that (a) what happens to me is my own doing, or (b) sometimes I feel that I don't have enough control over the direction my life is taking." The subject is to choose one of the two alternatives.

### 3. Subjects

During a lecture period in a course on the Psychology of Adjustment, 285 undergraduate students (177 females, 108 males) were given a battery of tests. Among these were the I-E Scale and the Orientation Inventory.

The Orientation Inventory (ORI) was scored by Bass's procedure and each subject was classified as self-oriented (S), interaction oriented (I), or task-oriented (T). Subjects placed in each category were those who received scores above the 75th percentile in the category in question *and* scores below the 75th percentile for each of the remaining categories. Subjects who had two or three scores above the 75th percentile were classified as residuals (R), as were subjects who had no score above the 75th percentile. The cutting points for the three scales (males and females combined) were: S—31, I—27, T—35.

The Ss in the upper and the lower quartiles of the I-E Scale were classified as High and Low scorers, respectively (cutting points were 12 and 15).

Because both experiments involved *Personality Category*  $\times$  *Sex* designs, in each of the designs the groups were balanced for proportions of Ss receiving the various category placements on the alternate test. In the design for the ORI, each of six groups contained 10 subjects. Three subjects had high scores on the I-E test, two subjects had low scores, and five subjects had intermediate scores. In the design for the I-E test, each of four groups contained nine subjects. One subject was self-oriented, three subjects were interaction-oriented, and five subjects were residuals. Such distributions were forced by the restrictions of the total sample. For example, very few females were classified as task-oriented (T). Data were available from a total of 94 Ss participating in the 10 groups of the two designs.<sup>1</sup>

After students had been classified, two undergraduate female assistants selected names randomly within the design categories and telephoned subjects

<sup>1</sup> The Low I-E-Female group and the T-Male group lost one S each because of unavailability of replacements for Ss failing to serve. Data for these groups were prorated.

for appointments.<sup>2</sup> In some categories subjects came from a relatively large pool of students and in others all students were used. Subjects were volunteers, but they received extra points toward their course grade.

#### 4. *Procedure and Apparatus*

The procedure used to measure *sr* was similar to that used in earlier experiments (4, 5, 8). The *S* was seated at a table on which rested a  $4' \times 3'$  hardboard screen to separate him from the experimenter who sat opposite. At eye level in the board was a  $4'' \times 6''$  window for the presentation of stimulus cards. A green jewel light beneath the window served as a positive reinforcing stimulus. The light was controlled by a microswitch that activated a two-second-delay relay.

The task for the subject was to learn a set of verbal discriminations. On each of 10 cards, four different nonsense syllables of medium association value were typed. The cards were presented in random order and *S* was required to spell one syllable on each card. On each card, a single syllable was designated as correct. Each of the four positions in which a syllable could appear was designated as correct on two, or three, of the 10 cards. For each presentation, a different random order of the cards was used. The presentations were repeated until the subject reached a criterion of 50 per cent correct. The flashing of the green light indicated a correct response.

After reaching the criterion, the subject received six blocks of *sr* trials. For these trials, he was given the microswitch that controlled the green light and was told to continue to select syllables as before and to press the switch whenever he was confident that he had chosen a correct syllable. The experimenter continued to present the stimulus cards and to record responses.

### C. RESULTS

#### 1. *Orientation Inventory*

Analysis of variance on trials to criterion indicates that the six groups did not differ significantly in rate of learning (highest  $F$  was 2.20;  $df = 2, 53$ ;  $p > .10$ ). The primary analysis on the *sr* phase of the procedure compared the groups on proportion of responses followed by an *sr* (converted to arc sines). The analysis of variance yielded a significant personality category and blocks interaction ( $F = 2.18$ ;  $df = 5, 265$ ;  $p < .05$ ). A test of each personality group for block trend shows that only task-oriented subjects increase in frequency of *sr* over trials ( $F = 3.36$ ;  $df = 5, 265$ ;  $p < .01$ ).

<sup>2</sup> Linda Wolfe and Barbara Selikoff assisted as experimenters.



The between-groups comparisons yield an  $F > 1.00$  only for the personality  $\times$  sex interactions ( $F = 2.46$ ;  $df = 2, 53$ ;  $p < .10$ ). Simple main-effect tests on this interaction indicate that there is a sex difference in frequency of sr only for the task-oriented subjects ( $F = 5.51$ ;  $df = 1, 53$ ;  $p < .05$ ). For the task-oriented group, the females took more self-reinforcement than the males.

Analysis of proportion of correct responses (converted to arc sines) during the sr test phase, shows a significant interaction of personality category and sex ( $F = 5.86$ ;  $df = 2, 53$ ;  $p < .01$ ). Simple main-effects tests show that the personality groups differ only for females ( $F = 4.76$ ;  $df = 2, 53$ ;  $p < .05$ ). The interaction-oriented females had fewer correct responses than either the self- or task-oriented females who did not differ significantly from each other (Tukey's gap test). The mean number of correct responses per block for the interaction females was four out of 10; for the self-oriented females, the mean was 5.4 out of 10; and for the task-oriented females it was 5.4 out of 10. In terms of sex differences in each of the personality groups, the simple main effect is significant only for the interaction-oriented group. In this group, females have fewer correct responses than males ( $F = 9.52$ ;  $df = 1, 53$ ;  $p < .01$ ). The mean per block for interaction-oriented males was 5.5 out of 10; for interaction-oriented females it was four out of 10.

## 2. I-E Scale by Sex Design

As in the preceding design, groups did not differ significantly in trials to criterion during the training phase. The highest  $F$  was 2.33 ( $df = 1, 31$ ;  $p > .10$ ). On proportion of responses followed by sr (converted to arc sines), analysis of variance shows a significant sex  $\times$  blocks interaction ( $F = 2.40$ ;  $df = 5, 155$ ;  $p < .05$ ) and a trend toward a sex difference in overall means ( $F = 3.01$ ;  $df = 1, 31$ ;  $p < .10$ ). In this design, male subjects increased significantly in frequency of sr over blocks ( $F = 2.27$ ;  $df = 5, 155$ ;  $p < .05$ ); while females showed no change ( $F = 1.07$ ;  $df = 5, 155$ ;  $p > .10$ ). The mean frequency of sr per block was 6.4 out of 10 for males and five out of 10 for females. Of the other comparisons on frequency of sr in the second experiment, only the triple interaction of personality, sex, and blocks yielded an  $F > 1.00$  ( $F = 1.73$ ;  $df = 5, 155$ ;  $p > .10$ ).

The analysis of variance on proportion of correct responses (converted to arc sines) yields a significant interaction of personality category and blocks ( $F = 2.60$ ;  $df = 5, 155$ ;  $p < .05$ ). Subjects who score low on the I-E scale (i.e., subjects with internal control of reinforcement) tend to increase

in frequency of correct responses over trials, while subjects who score high on the scale tend to decrease in correct responses over trials.

#### D. DISCUSSION

In previous studies (5, 8), Kanfer and the present author have found that frequency of sr tends to increase over trials. The present findings show that such increase is related to a personality variable measured by the Orientation Inventory. Only task-oriented individuals show increasing confidence in mastery of the task. By Bass's definition, self- and interaction-oriented individuals are more dependent on external or social rewards and are less concerned with accomplishment of the task set before them.

Sex difference in increase of sr was found in only one of two experiments. The difference in the two sets of data lies in the rate of self-reinforcement for females. They show a higher rate in the Orientation Inventory experiment than they do in the I-E design. Despite the fact that it was possible to balance the two designs for distribution as to sex and the categories on the alternate personality test, it was not possible to balance the same sex groups across the two designs. While males and females within one design are equated for personality distribution on both tests, the female group from one design was not equated with the group of females in the other design. For example, the female group in the I-E design contained 10 subjects classified as R on the Orientation Inventory; in the Orientation Inventory design, no Residual subjects were used. Therefore, conclusions regarding sex differences on rate of sr must be limited to the distribution of personality subgroups used in the I-E design.

As in earlier studies (5) it can be seen that an independent variable may affect rate of sr differently from rate of correct responses. There does not seem to be a clear relationship, such that an increase in sr causes poorer retention of correct responses. Earlier findings indicate that an increase in strength of correct responses causes an increase in sr, but the effect on correct responses of a manipulation of sr has not been clearly established. In the present study, the independent variable cannot be specified in terms of a sequence of effects; change in sr affecting correct responses, or *vice versa*. In the work with the personality variables, it was possible only to show similar or contrasting effects on the different dependent measures.

#### REFERENCES

1. BASS, B. M. *Leadership, Psychology and Organizational Behavior*. New York: Harper, 1960.

2. ———. The Orientation Inventory. Palo Alto: Consulting Psychologists Press, 1962.
3. DUNTEMAN, G., & BASS, B. M. Supervisory success and engineering assignment associated with Orientation Inventory Scores. *Person. Psychol.*, in press.
4. KANFER, F. H., & MARSTON, A. R. Conditioning of self-reinforcing response: An analogue to self-confidence training. *Psychol. Rep.*, 1963, **13**, 63-70.
5. ———. Determinants of self-reinforcement in human learning. *J. Exper. Psychol.*, 1963, **66**, 245-254.
6. KANFER, F. H., BASS, B. M., & GUYETTE, I. Dyadic speech patterns, orientation and social reinforcement. *J. Consult. Psychol.*, 1963, **27**, 199-205.
7. MARSTON, A. R. Tasks structure and the incidence of self-reinforcement. *Psychol. Rep.*, in press.
8. MARSTON, A. R., & KANFER, F. H. Human Reinforcement: Experimenter and subject controlled. *J. Exper. Psychol.*, 1963, **66**, 91-94.
9. ROTTER, J. B., LIVERANT, S., & CROWNE, D. P. The growth and extinction of expectancies in chance controlled and skilled tasks. *J. of Psychol.*, 1961, **52**, 161-177.
10. ROTTER, J. B., SEEMAN, M. R., & LIVERANT, S. Internal vs. external control of reinforcement: A major variable in behavior theory. In N. F. Washburne (Ed.), *Decisions, Values, and Groups* (Vol. 2). London: Pergamon Press, in press.
11. SKINNER, B. F. *Science and Human Behavior*. New York: Macmillan, 1953.

*Department of Psychology*  
*University of Wisconsin*  
*600 N. Park Street*  
*Madison 6, Wisconsin*

## ATTITUDES TOWARD CIVIL LIBERTIES AMONG JAPANESE AND AMERICAN UNIVERSITY STUDENTS\*<sup>1</sup>

*Department of Psychology, University of Maryland*

---

ELLIOTT MCGINNIES

---

### A. INTRODUCTION

Teenage attitudes toward a number of social issues, including civil liberties, have been studied in the United States by Remmers and Radler (3) and in Japan by Kato (2). Among other findings, Remmers and Radler reported that nearly half of America's teenagers are not committed to freedom of the press; over half approve censorship of books, newspapers, and magazines; and one-third believe that persons should be forced to testify against themselves if necessary. Another one-third do not believe in freedom of speech, and 26 per cent think that the police should be allowed to search a person or his home without a warrant. An interesting feature of these results, of course, is the fact that some of the expressed attitudes are in direct contradiction to the principles embodied in the Bill of Rights.

In contrast, using statements adapted from those employed by Remmers and Radler, Kato found that three-fourths of his sample of Japanese high-school students favored complete freedom of the press, nearly three-fourths believed that a person should not be compelled to testify against himself, and about 80 per cent were opposed to search by police without a warrant and to police wiretapping. Japanese college students tended to express similar opinions, except in the matter of censorship, which was favored by more than half of the high-school students, but by less than half of the college sample. Ninety-seven per cent of both groups of students thought that the government should not prohibit communists from making public speeches. Kato concludes that, in comparison with American teenagers, Japanese adolescents expressed more

---

\* Received in the Editorial Office on March 23, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> A number of individuals cooperated in securing the data for this report. Acknowledgement is gratefully made to Dr. F. H. Aiba, Dr. Y. Tanaka, Dr. G. Naruse, Dr. F. Nothman, and Dr. R. Lana. Miss Donna Bedgood did most of the tabulation and analysis of the data, Dr. T. G. Andrews provided statistical advice, and Nancy Kaye Sizer helped edit the manuscript. The project was supported jointly by the General Research Board of the University of Maryland and by the Group Psychology Branch, Office of Naval Research, under Contract Nour 3700 (01), NR 171-250 to the Institute for Behavioral Research.



democratic political attitudes; that is, more concern with people's rights and freedoms.

In view of Japan's relatively recent adoption of a Constitution modeled after that of the United States, these findings might strike one as paradoxical. Do democratic ideals tarnish with time, or do some indigenous qualities of Japanese youth predispose them to a closer affinity with certain democratic principles than is true of their American counterparts? The implications of answers to these questions suggest that closer inspection be made of attitudes toward civil liberties in the two countries, before attempting to explain the apparent differences. In the present study, a set of 12 items bearing directly upon civil liberties was administered to over 600 American and Japanese students at six different universities. Comparisons were made between students at universities in the same country as well as between students in the two samples as a whole.

## B. METHOD

As part of its promotional literature, the American Civil Liberties Union has devised a set of statements dealing with matters of concern to it. The respondent may agree or disagree with each item, and his total score indicates whether or not he is basically in sympathy with the stand on civil liberties taken by the ACLU. As expected, not all of the items on the ACLU questionnaire were appropriate for use with Japanese students; therefore, the statements were reviewed with two Japanese consultants. As a result, some of the statements were discarded and others were rewritten.

Twelve statements were adopted for use in assessing cross-national attitudes toward civil liberties in Japan and in the United States. Seven items were stated in a manner consistent with ACLU's attitude toward civil liberties, and five items expressed convictions inconsistent with ACLU's policies. A five-point rating scale accompanied each statement, with the alternatives labeled "strongly agree," "agree," "undecided," "disagree," and "strongly disagree." Both Japanese and English versions were prepared.

Responses to the items were obtained in June, 1962, from 299 students attending three Japanese universities and, in the spring of 1963, from 307 students at three universities in the eastern United States. Students at Waseda University, Tokyo University, and Kyushu University comprised the Japanese sample. The United States sample was composed of students at the University of Maryland, The American University, and Alfred University. Each group of universities contained two state-supported and one private institution.

No attempt was made to secure representative samples of respondents, as



this would have entailed more time and resources than were available. The questionnaires were given to psychology professors at the universities, who then distributed them to the students in their introductory classes. Because the students were drawn from a number of departments and ranged from freshmen to seniors, they are probably not grossly unrepresentative of the student bodies from which they were drawn. At no time did the author appear before the Japanese classes, so that the Japanese students did not associate the project with a foreign investigator. In all cases, the respondents were allowed to remain anonymous.

### C. RESULTS

The 12 statements dealing with civil liberties were not presumed to constitute a scale of attitudes toward civil liberties. Responses to each item, therefore, were analyzed separately. Two questions were asked of the data: (a) Were the students in each country sufficiently homogeneous to justify using their combined responses as an index of attitude on an issue? (b) Did the Japanese and American University students sampled hold significantly different attitudes on the issues studied?

To answer the first of these questions, the responses to each statement were subjected to chi-square analysis to determine whether university affiliation was independent of the distribution of individuals on the rating scale. Each chi square was thus computed from a  $3 \times 5$  array in which the rows were the three universities and the columns were the five response categories. The mean scale positions characterizing students in each of the universities were also computed, as additional measures of homogeneity. Cross-national differences were detected by chi-square calculations for the total samples of Japanese and American students for each of the 12 items, using  $2 \times 5$  contingency tables.

The tests for homogeneity of the Japanese and the American samples disclosed some discrepancies between universities. However, these reflected differences in the patterning rather than in the direction of the responses. Seven of the chi squares showed significant differences among the Japanese universities, and four revealed differences among the American universities. For example, in answer to the statement "police officials should have the right to listen in on private phone conversations," the mean scale positions for students at the three Japanese universities were 1.4, 1.5, and 1.8. Each of these means is in the direction of disagreement with the statement, although the less negative stand taken by one group of students led to a significant chi-square value. In response to the statement "police and other censors should be allowed to ban

books and movies that they consider to be 'indecent' or 'obscene,'" students at the three universities in the United States achieved mean scale positions of 3.9, 3.9, and 3.6. A significant chi square reflected the fact that one of the groups was less inclined to agree with the statement than were the other two.

Because there were no reversals in the direction of responses to any of the items among the Japanese or among the American student groups, the three samples in each country were combined for purposes of cross-national comparison. Due to the unequal sizes of the groups, percentages are shown rather than the raw frequencies from which the chi squares<sup>2</sup> were computed.

#### ISSUE NO. 1

GOVERNMENT EMPLOYEES ACCUSED OF DISLOYALTY SHOULD HAVE  
THE RIGHT TO KNOW THE SOURCES OF INFORMATION AGAINST  
THEM AND TO CROSS-EXAMINE THEIR ACCUSERS\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	4	9	18	49	20
U. S.	1	6	6	42	45

\* Chi square = 58,  $p < .001$ .

More of the Japanese students than of the American students were undecided about Issue No. 1, and the per cent of the Japanese who agreed strongly was decidedly less than the per cent of the Americans who agreed strongly. Otherwise, the trends were similar.

#### ISSUE NO. 2

POLICE AND OTHER CENSORS SHOULD BE ALLOWED TO BAN BOOKS AND  
MOVIES THAT THEY CONSIDER TO BE "INDECENT OR OBSCENE"

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	14	25	22	31	8
U. S.	5	18	9	33	35

\* Chi square = 87,  $p < .001$ .

Differences between the two groups of students were more pronounced on Issue No. 2 than on Issue No. 1, with the Americans favoring censorship to a much greater extent than the Japanese, nearly one-fourth of whom were undecided.

On Issue No. 3, direction of attitude was the same in the two groups, and the significant difference derived from the greater tendency of the American students to be egalitarian with respect to employment practices.

<sup>2</sup> For reporting purposes, chi-square values have been rounded to whole numbers.

## ISSUE NO. 3

PERSONAL ABILITY ALONE SHOULD DETERMINE EMPLOYMENT, REGARDLESS OF  
THE APPLICANT'S RACE, RELIGION, OR NATIONAL ORIGIN\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	0	8	14	39	39
U. S.	2	4	4	25	65

\* Chi square = 53,  $p < .001$ .

## ISSUE NO. 4

THE TEACHING OF SECTARIAN RELIGION SHOULD BE PERMITTED IN PUBLIC SCHOOLS\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	52	31	11	5	1
U. S.	3	13	12	34	38

\* Chi square = 278,  $p < .001$ .

Issue No. 4 resulted in a dramatic reversal of the two distributions, with the Japanese students rejecting the notion of religious instruction in public schools to an even greater extent than the American students favored it.

Issue No. 5 was the only issue for which the distributions of replies did not differ significantly between the Japanese and the American students. Both

## ISSUE NO. 5

REPRESENTATIVES OF THE EXTREME LEFT AS WELL AS THE EXTREME RIGHT SHOULD  
HAVE THE SAME PRIVILEGE OF MAKING PUBLIC SPEECHES AS DO  
OTHER POLITICAL LEADERS\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	4	10	16	46	24
U. S.	3	10	10	50	27

\* Chi square = 7,  $p > .05$ .

## ISSUE NO. 6

ANY PRIVATE INDIVIDUAL SHOULD HAVE THE RIGHT TO CRITICIZE ANY GOVERNMENT  
OR GOVERNMENT OFFICIAL ANYWHERE IN THE WORLD\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	0	8	15	38	39
U. S.	3	14	6	45	32

\* Chi square = 24,  $p < .001$ .

national groups tended to agree with the assertion, but with a sharp drop at the extreme position.

Although chi square was significant for Issue No. 6, the difference arose

from the tendency of the Japanese students to be undecided and from the tendency of the American students to disagree. Actually, 77 per cent of each group agreed with the statement.

ISSUE NO. 7  
POLICE OFFICIALS SHOULD HAVE THE RIGHT TO LISTEN IN ON  
PRIVATE PHONE CONVERSATIONS\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	63	24	9	3	1
U. S.	3	12	10	38	37

\* Chi square = 368,  $p < .001$ .

On Issue No. 7, the two distributions were completely opposite in direction, with the Japanese students even more united in opposition to police "eavesdropping" than the American students were in acceptance of this activity.

ISSUE NO. 8  
EMPLOYERS SHOULD BE PERMITTED TO STATE THEIR VIEWS  
REGARDING LABOR UNIONS TO THEIR WORKERS\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	2	4	12	64	18
U. S.	1	4	5	58	32

\* Chi square = 22,  $p < .001$ .

For Issue No. 8, the two distributions were strikingly similar. The significant chi square derived mainly from the fact that a higher proportion of American students agreed strongly with this position, while more of the Japanese students disagreed or were undecided. This result may reflect a

ISSUE NO. 9  
MOVIES, PLAYS, AND BOOKS SHOULD BE SUPPRESSED IF THEY PRESENT AN OFFENSIVE  
CHARACTERIZATION OF A PARTICULAR RACIAL OR RELIGIOUS GROUP\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	10	41	20	24	5
U. S.	7	16	11	46	20

\* Chi square = 93,  $p < .001$ .

more strongly pro-management disposition on the part of the American sample than on the part of the Japanese sample.

Issue No. 9 provided an acid test of one's commitment to freedom of

speech and found significantly greater support among the American students than among the Japanese students. The Japanese students were apparently more willing than the American students to extend freedom of expression even to those individuals who use it to malign minorities.

ISSUE NO. 10  
A PERSON WHO CLAIMS THE PRIVILEGE AGAINST SELF-INCRIMINATION WHEN  
ASKED IF HE IS A COMMUNIST IS NOT NECESSARILY A  
COMMUNIST SYMPATHIZER\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	1	3	12	55	29
U. S.	4	11	16	51	18

\* Chi square = 32,  $p < .001$ .

Both the Japanese and the American students showed the same general trends of response to Issue No. 10, but the American students did not support the statement as strongly as did their Japanese counterparts. In short, the Americans were more inclined than the Japanese to the suspicion that a person who refuses to make a possibly self-incriminating statement may indeed be guilty as charged.

ISSUE NO. 11  
TESTS OF GOVERNMENT EMPLOYEES' LOYALTY SHOULD BE CONFINED TO SENSITIVE  
POSITIONS INVOLVING MILITARY, ATOMIC, OR INTERNATIONAL AFFAIRS\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	24	32	37	6	1
U. S.	13	38	17	23	9

\* Chi square = 81,  $p < .001$ .

The Japanese students were less inclined than the American students to favor the point of view of Issue No. 11, with 56 per cent of the former group apparently convinced that government employees should be proved loyal in a broader sense. Many of the Japanese, however, were undecided, and nearly one-third of the Americans were content to confine loyalty investigations to areas involving national security. The American students were bimodally distributed on the issue, a finding consistent with the controversy that has surrounded the problem for some time.

On Issue No. 12, a bimodal distribution characterized the American students, although more of them than of the Japanese students were inclined to agree with the statement. The Japanese students, in contrast, were decidedly opposed to a loyalty oath for teachers.



ISSUE NO. 12  
PUBLIC-SCHOOL AND COLLEGE TEACHERS SHOULD BE REQUIRED TO  
SIGN A NON-COMMUNIST LOYALTY OATH\*

Country	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Japan	47	36	14	2	1
U. S.	17	25	13	30	15

\* Chi square = 153,  $p < .001$ .

### D. DISCUSSION

In addition to determining whether response patterns differed between the Japanese and the American samples, on each item the mean scale position of the two groups of students was computed. This was done by scoring 5 for "strongly agree" if the statement expressed a stand taken by the ACLU, and -5 for "strongly disagree" if the statement negated that position. With this system of weighting, it was found that both groups, on the average, defended the following principles: the individual's right of cross-examination in loyalty cases, personal ability as the sole criterion for employment, freedom of expression for extreme "leftists" as well as for extreme "rightists," the right of an individual to criticize his government, the rights of employers to present their own viewpoints to their workers, and the right of a person to invoke the privilege against self-incrimination.

The Japanese university students tended to disapprove of religious instruction in the public schools, of police wiretapping, of suppression of ideas that might offend some particular group, and of the requirement for a loyalty oath from teachers. The American students acceded to the first three of these infringements on civil liberties and, in addition, agreed to censorship of books and movies by the police. The only issue on which the Japanese students could be said to have taken a less libertarian viewpoint than the American students was in regard to tests of government employees' loyalty, which tests they thought should extend beyond sensitive positions involving international and military matters. The Americans were divided on the question of loyalty oaths for public-school and college teachers, and the Japanese were divided on the question of censorship.

### E. INTERPRETATION

Our results, while supporting earlier indications of a greater concern by Japanese youth than by American youth with civil liberties, are puzzling—in view of Japan's recent history of political authoritarianism and regimentation. While the American students sampled in this study appeared to have lost

sight of some of the liberties protected by the Bill of Rights, the Japanese students gave evidence of having adopted these same guarantees with an enthusiasm that seems inconsistent with their cultural and political heritage. This finding clashes with the prognostications of certain other studies of Japanese youth. For instance, Stoetzel (4), in his study of the attitudes of youth in postwar Japan, expressed misgivings concerning the viability of an egalitarian philosophy among the Japanese and suggested that a fundamentally authoritarian outlook underlay much of the apparent interest of youth in democratic principles.

Actually, it may be fallacious to assume that Japanese society prior to World War II contained none of the elements upon which a democratic ideology might successfully be grafted. Benedict (1) recognized in the immediate postwar years that "crass authoritarianism" was not the "problem" in Japanese culture. She implied further that the peculiar ethic of the Japanese would allow them to abandon their current political and social course and to adopt the constitutional reforms pressed upon them by the occupation authorities.

Subsequent developments seem to support Benedict's predictions. Certainly there are many indications, above and beyond those contained in scholarly reports, of an active concern on the part of Japanese youth with those freedoms of thought and action that are given expression both in their Constitution and in the United States Bill of Rights. These convictions are shared to an extent with American students, as the findings reported here indicate, but the greater balance of student opinion supporting civil liberties at the present time seems to lie with the Japanese.

Perhaps it should be pointed out that many Japanese students who are politically liberal—even radical—during their college years develop into conservative businessmen after graduation. This metamorphosis does not exclude leaders of the Zengakuren, the ultra-leftist student organization. The implication (one that has also been offered by several Japanese professors) is that libertarian attitudes may characterize a student "role," even to the extent of being expressed dramatically in public demonstrations. Only a longitudinal study of attitudes toward civil liberties, involving individuals who have graduated several years prior to the study, will indicate whether Japanese students hold these views with any tenacity or permanence.

Finally, it should be emphasized that extrapolation beyond the groups studied should be made with caution. The present findings should be useful in devising a more intensive analysis than this one of student attitudes in the

two countries, by use of larger and more carefully controlled samples than these.

Further interpretation of cultural differences in libertarianism must wait upon detailed analysis of the dimensions of the concept of "civil liberties" and of the issues that are included in its scope. The data suggest that several factors may be operative, and it is perhaps upon the investigation of these more specific attitudes that an adequate discussion of cross-national differences must rest.

#### REFERENCES

1. BENEDICT, R. *The Chrysanthemum and the Sword*. Boston: Houghton Mifflin, 1946.
2. KATO, T. Political attitudes of Japanese adolescents. *Psychologia*, 1961, 4, 198-200.
3. REMMERS, H. H., & RADLER, D. H. *The American Teenager*. New York: Bobbs-Merrill, 1957.
4. STOETZEL, J. *Without the Chrysanthemum and the Sword*. Boston: Houghton Mifflin, 1946.

*Department of Psychology*  
*University of Maryland*  
*College Park, Maryland*



## THE CALIFORNIA PSYCHOLOGICAL INVENTORY AS A PREDICTOR OF YIELDING BEHAVIOR IN WOMEN\*

*Institute of Child Development, University of Minnesota*

---

FRANK B. W. HARPER

---

### A. INTRODUCTION

Tuddenham (3) reported a study of the personality correlates of yielding to a distorted norm, in two samples of women—one a college sample and the other an older women's sample of more heterogeneous background. Comparing scores on the 18 scales of the California Psychological Inventory (1) with yielding scores, he found that, in the older sample, three scales were significantly related to yielding. These were Capacity for Status, Dominance, and Sociability. In the college sample, four scales were significantly related to yielding. These were Capacity for Status, Flexibility, Achievement via Independence, and Social Presence. Although Tuddenham did not conduct a predictive study and his results are subject to cross validation, the findings are congruent with the theoretical rationale for each of the scales, as advanced by Gough (1, pp. 12-13). For example, high scorers on Capacity for Status are described as ambitious, ascendant, and self-seeking, while low scorers are described as conventional, stereotyped in thinking, and uneasy in unfamiliar social situations. Under the assumption that the qualities required to maintain independence in the face of the false answers of the group are those qualities possessed by high scorers, and that the qualities leading to yielding are those qualities possessed by low scorers, it would be a rational prediction that scores on the Capacity for Status scale should be negatively related to yielding. As such, the obtained significant correlation could be construed as evidence for the construct validity of the scale. A similar argument could be made for interpreting the results of the other scales as evidence for their construct validity, except for the inconsistencies in the findings across the two samples. Only one scale, Capacity for Status, was significant for both samples. The other scales were significant on only one sample. The findings may be viewed, therefore, as only tentative and as requiring replication.

The present investigation is a partial replication of Tuddenham's study, in that yielding is measured under similar conditions, on the same apparatus, and

---

\* Received in the Editorial Office on March 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.



compared to scores on the California Psychological Inventory. The subject population, however, differs in that a larger sample of women is used which resembles Tuddenham's college population more than his older women's sample. A number of predictions are made concerning the relation between certain CPI scales, based in part on Tuddenham's findings and on the theoretical rationales and empirical descriptions of high and low scorers on the scales, given by Gough.

### B. HYPOTHESES

It is predicted that scores on the individual scales—Capacity for Status, Dominance, Achievement via Independence, Flexibility, Sociability, and Social Presence—are negatively related to yielding scores in women. The qualities measured by these scales are summarized as follows: High scorers are described as having qualities of leadership, independence, self-reliance, self-confidence, ambition, persuasiveness and forcefulness. Low scorers are described as retiring, inhibited, avoiding of situations of tension and decision, lacking in self-confidence, submissive and compliant before authority. It is assumed that subjects possessing the qualities ascribed to high scorers are more able to resist the distorted group norm than are subjects possessing the more negative qualities ascribed to low scorers.

### C. SUBJECTS

The subjects were 135 student nurses from a private school of nursing, ranging in age from 17 to 22, with a median age of 19. Their mean *IQ* was 114, with a standard deviation of 10. Although the subjects had a common aim in studying for certification as nurses, they were a varied group in their background and ability.

### D. PROCEDURE

The subjects were tested in groups of five in the experimental setting described fully elsewhere by Tuddenham (2). The amount of yielding to a distorted group norm was calculated for each subject over a series of trials. In addition, each student filled out the California Psychological Inventory at another time and place.

### E. RESULTS

The product-moment correlations between yielding scores and each of the six CPI scores are given in Table 1. Only two scales, Capacity for Status and Achievement via Independence, had correlations that significantly ex-

TABLE 1  
MEANS, STANDARD DEVIATIONS, RANGES OF SCORING, AND CORRELATIONS OF  
SIX CPI SCALES WITH YIELDING

Scale	Mean	SD	Range of scoring	Correlation with yielding
Dominance	27.00	6.19	12-39	-.10
Capacity for Status	20.04	3.94	9-29	-.17**
Sociability	25.90	4.67	14-35	-.06
Social Presence	34.92	5.49	19-46	-.03
Achievement via Independence	19.93	3.69	8-28	-.19**
Flexibility	9.90	3.45	2-18	-.10

\*\*  $p < .025$ .

ceeded zero. The values of  $r$  for both scales are, however, so low that they account for only a small fraction of the total variance. A check on the variability and the range of scoring for each scale was made to test the possibility that the low correlations were a function of unusual restriction in the range of scores for this sample. The means, standard deviations, and ranges of scores are shown in Table 1. The distributions of scores for each scale are comparable to those for the standardization populations reported by Gough (2, p. 35), and so the low correlations do not arise from restriction of the sample.

A test for curvilinearity of relationships was carried out on each bivariate distribution, but the scatter plots were consistent with the assumptions of linearity of regression. The obtained correlations are therefore interpreted as evidence for a low or zero relationship between the CPI scales and yielding.

## F. DISCUSSION

Of the two scales that were significantly related to yielding, one scale, Capacity for Status, was also related to yielding in both of Tuddenham's samples. It appears, therefore, as if this scale does have validity for predicting yielding behavior in women, and, because the results are derivable from the theory of the scale, the obtained correlation may be interpreted as a construct-validity coefficient. The same interpretation may be applied to the finding for the Achievement via Independence scale. This scale was also correlated with yielding in Tuddenham's college sample that was comparable to the present sample. However, the proportion of yielding variance which is accounted for by scores on either the Capacity for Status scale or the Achievement via Independence scale is so small as to suggest that knowing only these scores would not help much in predicting whether women would

yield or not. The major personality correlates of yielding in women are evidently not those measured by these scales of the CPI. Comparison of the correlations between all CPI scales and yielding scores showed that two other scales were significantly related to yielding. These were Tolerance, which had a correlation of  $-.20$  with yielding, and Good Impression, which had a correlation of  $.19$  with yielding. Both were significantly different from zero at the  $.05$  level, but again the correlations were so low as to make their predictive values limited.

### G. SUMMARY

Predictions concerning the relation between six scales of the California Psychological Inventory and yielding to a distorted norm, for a sample of 135 women, were made on the basis of the theoretical rationales for the scales, plus other research findings. Two scales, Capacity for Status and Achievement via Independence, were significantly negatively related to yielding scores, but the correlations were so low as to account for only a small fraction of the total variance. Yielding behavior in women appears to arise from personality variables other than those measured by the California Psychological Inventory.

### REFERENCES

1. GOUGH, H. G. Manual of the California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
2. TUDDENHAM, R. D. The influence of a distorted group norm upon individual judgment. *J. of Psychol.*, 1958, **46**, 227-241.
3. ———. Correlates of yielding to a distorted group norm. *J. Personal.*, 1959, **27**, 272-284.

*Institute of Child Development  
University of Minnesota  
Minneapolis 14, Minnesota*

## THE EFFECT OF ANXIETY ON THE PERFORMANCE AND ATTITUDES OF AUTHORITARIANS IN A SMALL GROUP SITUATION\*<sup>1</sup>

*Department of Psychology, Radford College*

---

CARROLL E. SMITH

---

### A. INTRODUCTION

A review of the literature reveals (a) that attempts to change the stereotyped attitudes of authoritarian persons, through the use of argument and exhortation, have been rather uniformly unsuccessful (7, 8, 15, 22); (b) that extensive, but conventional, contact with minority group members probably reinforces—rather than reduces—prejudiced attitudes (2, 3, 9); and (c) that the emotional component may remain active in the behavior of the prejudiced individual even though he has intellectually accepted the logic of tolerant ideas (10, 17).

Other studies have reported attitude changes when some special interpersonal interaction between the authoritarian individual and members of the devalued minority is involved (3, 11, 19, 22). This is particularly the case if there is a shared emotional experience or a response to common dangers (2, 12, 21). Related experiments by Sherif have shown that tension reduction between mutually hostile groups can be achieved by the necessity of cooperation on a task of concern to both groups (17).

In general, these studies point to the conclusion that attitudinal changes must be mediated by some sort of emotional experience. Under the influence of suitable motivation, the emotional experience would be expected to support any cognitive reorganization that would be consistent with the emotional aspect of the interpersonal experience.

Festinger's study of cognitive dissonance (9) would seem to lend theoretical support for study in this area. From Festinger's theory, it would be predicted that the authoritarian person who had been placed in a position in which there was close association in a positive emotional climate (e.g., facing danger together, forced cooperation, etc.) would feel a dissonance between his experiences and his attitudes and opinions about the minority group in question.

---

\* Received in the Editorial Office on March 27, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This investigation was supported by Public Health Service Research Grant MH 07299-01, from the Institute of Mental Health.

This state of cognitive dissonance would provide the motivation for a change in attitude that would serve to reduce the dissonance and the associated motivational factors. In a similar situation, the nonauthoritarian person would find his experiences consonant with his cognitions about the minority person, and there would be no drive to reduce.

Anxiety is a second construct that has received considerable recent use as a variable in studies of small groups. On an operational level, differences between various studies can be attributed to the considerable difference between the various anxiety-measuring techniques: clinical, physiological, or inventories. In addition, differences in conceptualization (anxiety as a trait or interpsychic structure, anxiety as a dispositional term or low threshold for anxious arousal, etc.), plus the drive-increasing properties of activated anxiety, have made it a difficult concept to use with any assurance. The difficulty is compounded by the fact that the studies in group participation in which anxiety has been used as an experimental variable have shown inconsistent and contradictory results. A good summary on this point is that prepared by Sarason (16).

The theory of anxiety that was used in this study is that hypothesized by Spence and Taylor [see Spence (20)], and the instrument used for ranking the subjects on this variable was Taylor's Manifest Anxiety Scale (MAS). When applied to the authoritarian subject in a dissonant situation with a minority person, the following theoretical position was assumed: the authoritarians who are placed in a situation in which there is a dissonance between attitudes and behavior will be motivated toward a reduction of the hostile feelings toward minority figures. If the situation should involve an emotional experience, persons with a low-anxiety threshold would be expected to show a more extreme modification of attitude due to the drive-increasing effect of the aroused anxiety (14).

The preceding assumptions would apply in the case of uncomplicated interaction with the minority group. However, in the case of a direct conflict between two possible responses—one consistent with prejudiced views and one conflicting with these views—we would expect a differential effect from the anxiety level as a function of the well-established disorganization of perceptions and performances under anxiety-arousing conditions. Thus, we might expect anxious authoritarians, in a conflict situation in which a choice is possible, to be more strongly motivated than nonanxious authoritarians to seek the security of old associations and attitudes—even if this seeking entails a denial of an experience with a minority group. This action should also (as a dissonance-reducing technique) reinforce and strengthen the anxious authoritarian's attitude toward the minority group.



To test our ideas, a small group experiment modeled after Asch's line-judging technique (4) was chosen. Asch has found that the provision of a coached *partner* for a naive subject not only increases the subject's resistance to conformity pressure, but also promotes a positive emotional reaction toward the person who has provided support against the majority. In the case of the nonauthoritarian (or democratic) subject, the outcome should be similar to that reported by Asch: namely, greater resistance to conformity and positive feeling for his *partner*.

For the authoritarian subject, there should be a dissonance between his perception of the line matching and his alliance with the minority person in opposition to the majority. In this case, the expectation would be that the authoritarian subject would *not* feel the emotional support of a partner, but would be more strongly motivated toward a conformity that would escape association with the minority person. If the authoritarian subject is anxiety prone, the experience might cause his anxiety threshold to be exceeded and cause him to be even more strongly motivated toward conformity. The disorganizing effect of heightened anxiety upon performance would facilitate conformity, even though it would mean greater distortion in the perception of the objective stimulus situation.

Following the group experience, it would be expected that the authoritarian subject would attempt to reconcile his behavior and attitude. In the case of the anxious authoritarian (who should be less task oriented than his nonanxious counterpart), a high degree of yielding (*a*) to majority pressure in opposition to the provided minority *partner* and (*b*) to the objective task situation (plus the added drive of aroused anxiety) should result in increased negative feeling toward the minority *partner* and toward his group. Also, the increased drive should raise the gradient of generalization, and an increase in negative attitudes toward other groups should occur.

In the case of nonanxious authoritarians, more task orientation and better performance should result in a dissonance between the forced alliance with the minority *partner* and his devaluative attitudes. In such a situation, dissonance would be reduced by a more favorable attitude toward the minority *partner* and toward the group which he represents. Because this dissonance-reducing technique would reduce the associated drive, and because there would be no anxiety interaction, there would be no generalization of the more favorable attitude to other minority groups.

For democratic subjects (who are considered as a control), there should be no change in attitude because the entire situation would be attitude-irrelevant for them.

## B. HYPOTHESES

Applying the preceding theoretical development to the Asch line-judging situation, we made two sets of predictive hypotheses: those concerning behavior in the experiment and those relating to postexperimental attitudes.

### 1. *Experimental Behavior Hypothesis*

As a group, (a) authoritarian subjects will show a greater degree of conformity (higher yielding scores) to the majority than will democratic subjects; (b) anxious authoritarians will obtain higher yielding scores than will nonanxious authoritarians; and (c) there will be no difference in yielding scores between anxious and non-anxious subjects.

### 2. *Postexperimental Attitudinal Hypothesis*

As a group, (a) nonanxious authoritarians will show a more favorable attitude toward an experimental minority group on a postexperimental scale, as compared to a pre-experimental scale; (b) anxious authoritarians will be more severe in their evaluation of the minority group on a postexperimental scale, as compared to a pre-experimental scale; (c) the more favorable attitude of nonanxious authoritarians toward the specific minority group will not affect their judgment of other groups; (d) the increase in unfavorable attitudes toward the minority on the part of the anxious authoritarians will generalize, and they will show a more hostile attitude toward other groups (greater total score); and (e) among democratic subjects, there will be no significant difference in postexperimental attitudes for either the high-anxiety or the low-anxiety groups.

## C. EXPERIMENTAL PROCEDURE

### 1. *Subjects*

The subjects for the experiment were female freshman students entering college for the first time. The average age was 18 for all groups. The subjects were homogeneous with respect to College Entrance Examination Board (CEEb) scores, geographic distribution—and probably socioeconomic status, although this was not checked further than by an inspection of a personal-data sheet.

### 2. *Method*

The subjects were selected on the basis of a survey of student opinion among the freshman class at Radford College. The survey was comprehensive, covering attitudes toward the college and political, economic, and moral

questions. The sections were used to support the idea of a comprehensive survey and to make less noticeable the particular scale of interest in this study. Included in the survey were (a) the 30-item version of the California F Scale, (b) the Taylor Manifest Anxiety Scale, and (c) a student modification of the Bogardus Student Social Distance Scale (SSDS). The social-distance scale was scored by summing the scale values for each scale. By use of a scale of nine steps and 14 groups, it was possible to score between zero and 24, the higher score indicating greater social distance. (a), (b), and (c) were identified in the survey only as parts B, D, and G.

The experimental subjects were selected on the basis of extreme F-Scale and MAS scores to form four groups: Authoritarian, Anxious (HH); Authoritarian, Nonanxious (HL); Democratic, Anxious (LH); and Democratic, Nonanxious (LL). The original selection was 20 subjects per group. In the F-Scale selection of authoritarians, all subjects scored over 20; the democratic subjects all scored below 7 on the same scale. The same (apparent) degree of separation could not be achieved on the scale for Manifest Anxiety, the separation being below 19 for low subjects and above 20 for high subjects.

Due to the necessity of discarding some results because of the suspicion that the purpose of the experiment had become known to some of the subjects, the final number in each group (with their median scores on the two selection variables) was as given in Table 1.

TABLE 1  
MEDIAN AND RANGES ON THE F SCALE AND THE MANIFEST ANXIETY SCALE  
FOR THE FOUR EXPERIMENTAL GROUPS

Group	N	F Scale		Manifest Anxiety Scale	
		Median	Range	Median	Range
HH	20	24.5	20-26	26.0	21-35
HL	18	23.0	20-27	16.0	5-18
LH	19	5.5	4-7	25.0	21-31
LL	16	5.0	3-7	13.5	7-17

Significance tests indicate that the grouping (on the basis of within-group homogeneity and between-group separation) is satisfactory. There is the expected correlation between F-Scale scores and scores on the Social Distance Scale.

The before-mentioned line-judging situation of Asch was employed in the following manner:

1. The majority group of coached confederates consisted of four female students that might be described as "standard American" types. They were selected (on the basis of the impression of the experimenter) as having no special feature that would identify them with anything other than the average American girl.

2. The coached *partner* for the subject was a Chinese student. She had the same student status as the other confederates and the subjects. Her spoken English was fair.

3. In the experimental situation, all six participants arrived at the designated time. The experimenter explained the procedure as though it was the first time for all of the participants. By prior arrangement, the subject was seated at the end of the row of subjects, necessarily responding last. The minority-group *partner* was seated two seats away, being fourth to respond.

The line judging followed by placing a card, with a standard line at the left side and four choice lines at the right side, on a stand in front of the room. The group was asked to respond aloud with the number of the line chosen as matching the standard line (proceeding from the first majority member and ending with the subject). Each group member was asked to record (on a sheet of paper attached to a clipboard) her response and the number of persons agreeing with her choice. The papers given to the experimental confederates also included marginal notations of the line choices to make for each card. The experiment proceeded through 30 cards. Included in the 30 cards were 12 critical cards for which the majority unanimously agreed on a particular incorrect line and the minority *partner* of the subject chose the line that was a match for the standard line.

Because performance would be affected by the objective structure of the stimulus situation, each subject's score was determined by summing the number of times she yielded to the majority on the critical matches, weighted by the difference in the line chosen by the majority and the standard line. Thus, when the difference was less than 10 per cent, the yielding score was 1; if 10 to 20 per cent, the yielding score was 2; and over 20 per cent, 3. Four of the 12 items had each of these values (i.e., 1, 2, and 3), making it possible for a subject to score between zero and 24. After the experimental procedure, each subject was asked to repeat the Student Social Distance Scale, plus a similarly designed scale (called the Personal Social Distance Scale, PSDS) upon which persons other than the subject in the experimental situation were the focus of inquiry. The items of the PSDS scale were as follows:

1. Would accept her as a roommate.
2. Would accept her as a friend, but not as a roommate.



3. Would accept her as a member of my sorority or social club.
4. Would accept her as a resident of my dormitory.
5. Would accept her as a member of my academic club.
6. Would accept her as a member of my classes.
7. Would accept her as a student at my college.
8. Would prefer that she would not be permitted to attend my college.

Following the completion of the PSDS scales, the subjects were interviewed briefly to determine whether they had any knowledge of the purpose of the experiment. Because it is unlikely that the steps on the Personal Scale are comparable to those of the Student S-D Scale, a control group of 20 authoritarian subjects was selected. These subjects were run through the line-judging procedure with the experimental confederates. Instead of the standard procedure, each person selected whatever line seemed proper. The minority confederate was present, but was not allied with the subject. Each of the subjects completed the Personal Social Distance Scale.

#### D. RESULTS

Because the data do not meet the requirements for the use of parametric statistical tests, the results were analyzed by the use of procedures appropriate for ordinal, nonnormal data, as described by Siegal (18) or by McNemar (13).

TABLE 2  
COMPARISON OF YIELDING SCORES FOR THE FOUR EXPERIMENTAL GROUPS  
(Chi-square test for difference between medians)

Group	Median	Difference	Chi square	p
HH	10.5	3.0	3.88	< .05
HL	7.5			
LH	3.3			
LL	4.0	.7	.0079	> .90
Total				
Authoritarian	9.5	5.9	18.94	< .001
Democratic	3.6			

The yielding scores for the four groups of subjects are presented in Table 2. Table 2 shows that the three performance hypotheses for the experimental situation are in the predicted direction. The statistical test is the chi-square test for difference between medians.

Table 3 gives the statistical test for the differences in scores on the pre- and postexperimental administrations of the Student Social Distance Scale.



TABLE 3  
COMPARISON OF PRE- AND POSTEXPERIMENTAL ATTITUDES (MEDIAN) ON THE  
STUDENT SOCIAL DISTANCE SCALE  
(McNemar test for changes and Binomial test)

Group	Social Distance Scale		Chi square	<i>p</i>
	Pretest	Posttest		
HH	35.0	46.0**	4.166	< .05
HL	38.0	40.0	Binomial*	> .10
LH	9.0	10.0	Binomial*	> .50
LL	11.0	11.5	Binomial*	> .50

\* Binomial test was used when frequencies were too small to permit use of the McNemar test.

\*\* 44.0 when corrected.

In this analysis, the postexperimental distribution of the HH group was dropped two points to evaluate the net change in performance as compared to the HL group, for which the median was two points higher on the posttest as compared to the pretest. This net difference was then evaluated by the McNemar test for the significance of changes, as described by Siegel (18, pp. 63-67). The pre-experimental versus postexperimental difference is significant at the .05 level for the HH group.

The expected frequencies for the other groups are too small for use of the McNemar test. These differences were evaluated by the binomial test (18, pp. 36-42). All differences are small, are likely to occur as errors, and are considered nonsignificant.

Thus, the attitudinal hypotheses are supported in part: HH subjects (as predicted) become more negative in their evaluation of the groups listed on the Student Social Distance Scale. Neither the HH group nor any of the other groups show a significant amount of change on the individual scale representing the minority *partner* (not tabled).

Data for the two groups of authoritarian subjects on the Personal Social Distance Scale, as compared with the authoritarian control group, are given in Table 4. The control group was necessary in this comparison because there could be no pre-experimental use of this scale: the subjects were not acquainted with the persons to be rated until after the experiment. In this analysis, the ratings of the control group are used as the expected frequency for a chi-square analysis of the observed frequencies.

From Table 4, it can be seen that only the HH group shows a significantly more negative evaluation of the minority person than do the authoritarian controls. There is some change in the same direction for the HL group, but not enough to reach conventional standards of significance.

TABLE 4

COMPARISON OF ATTITUDES TOWARD THE MINORITY PARTNER BY THE TWO AUTHORITARIAN EXPERIMENTAL GROUPS AND THE AUTHORITARIAN CONTROL GROUP ON THE PERSONAL SOCIAL DISTANCE SCALE\*

Score	Control group and HH group**		Control group and HL group***	
	Expected frequency	Observed frequency	Expected frequency	Observed frequency
Above median	5	11	5.4	8
At median	9	7	8.1	6
Below median	6	2	4.5	4

\* Control group frequencies are used as expected frequencies for comparison with the behavior of the experimental group. The classification was condensed so that the expected frequencies would be sufficiently large.

\*\* Chi square = 10.30,  $p < .01$ .

\*\*\* Chi square = 3.30,  $p > .10$ .

### E. CONCLUSION

For the democratic subjects, the results support the prediction that the experimental procedure would not differentially affect the anxious and non-anxious subgroups. The alliance with the minority partner seemed to be acceptable and served to support the democratic subjects to the extent that they were able to resist the pressure toward conformity to the majority opinion. The absence of any changes in pre- and postexperimental attitudes was as expected, as the situation was consonant with the attitudes of these subjects.

In the case of the anxious authoritarian (HH) group, the conflict between the response that would correspond to visual reality and the response that would correspond to the subject's concept of social reality agrees with the conceptualization of the cognitively dissonant conflict situation. The acceptance of the majority opinion by this group of subjects was done (a) with the knowledge that the minority persons involved held a different opinion and (b) with the knowledge that the opinion held by the minority was the correct one (i.e., it was perceptually accurate). In this situation, there would be a residual dissonance that might be reduced by an increase in hostility toward the minority person in the experiment and the group represented. The summation of the cognitive-dissonance-produced drive and the anxiety-produced drive should raise the gradient of generalization to the extent that the increase in hostility should spread to other groups. Our results do show an overall rise in the social distance score for anxious authoritarian subjects. The observed change in the Chinese scale is, however, not sufficient to produce statistical significance (not tabled). Because the Chinese were scaled

high on the original ranking, a large change in score would go to the end of the scale and necessitate the exclusion of the group from the country. We find that such an extreme position is rarely endorsed by even the most extreme authoritarians, perhaps because of the extreme conflict with democratic social ideals.

The position of the minority *partner* is significantly more negative for anxious authoritarian subjects than for the authoritarian control subjects and the nonanxious authoritarian subjects. In this instance, the relatively high position of the minority *partner* as rated by the control group provides ample range for a more negative position by anxious authoritarians. This is, of course, the expected outcome of the dissonance-reducing attitude shift for the group. The fact that we did not find the expected amount of negative generalization may be attributed to the limited or artificial nature of this type of group and to the fact that the minority *partner* is something of a "special case" and is not perceived as truly representative of her membership group.

Although the performance of the nonanxious authoritarians in the experimental situation agrees with the predicted behavior, there was no noticeable change toward a more favorable attitude as a result of their greater agreement with the minority *partner*. This is, of course, the point of primary concern because of the desirability of promoting a more democratic point of view among prejudiced persons.

Without trying to explain away negative results with *post hoc* propositions, it does seem that there may be good reasons for our results.

The predicted outcome was that the result of alliance with the minority *partner* would lead to a reduction of dissonance through a more favorable evaluation of that person and of his group. As mentioned above, the particular minority person involved in the experiment had a special status. First of all, she was a student on campus, one of the few foreign students on campus and, therefore, not a typical representative of her group. Since it has been observed by others that a few special members of devalued minority groups may be given special status (i.e., the "pseudo-white" status of Negro celebrities, etc.), it is likely that our minority-group member would be seen to occupy a special position and that any experience with her would not extend to other members of her group. This idea is supported by the fact that the members of the control group of authoritarians were relatively accepting of this girl, finding her an acceptable dormitory resident.

A second factor is the general democratic atmosphere of the campus. All minority groups on campus make up a very small percentage of the student

body. They are too few minority persons to have any special clubs or organizations (foreign students' clubs, B'nai B'rith, etc.), and these persons are randomly distributed through the various dormitories and classes. This factor would tend to facilitate the assignment of a special, nonrepresentative status to these persons.

The performance of the anxious authoritarian subjects can be seen as being consistent with this situation. The experimental situation would have been more threatening to them. The hostility directed toward the minority *partner* as a dissonance-reducing technique should have been at a higher level than the countermotivation of the nonanxious authoritarians toward a more friendly attitude and should have produced greater change. Moreover, the fact that the minority *partner* already occupied a favorable social position as compared to her group would make it easier to move to a higher scale position than to a low one, in this situation. The increase in the hostile evaluation of the minority person is also consistent with the proposition of heightened drive.

An additional consideration is the possibility that the character organization of the authoritarian personality is such that any anxiety-arousing experience brings out a relatively nonspecific response of hostility (i.e., scapegoating). In such a case, it is possible that there would be an increase in anti-minority feeling among anxious authoritarians following any frustration or threat, regardless of the presence or absence of a minority member. The corollary of this idea might be that the authoritarian person lives in a more or less constant state of cognitive dissonance between his antiminority attitudes and his knowledge of the generally accepted democratic ideals of contemporary Americans. If this were then combined with a low threshold for anxious arousal, there would be a need for a continued search for reports or ideas about the minority group, which ideas would be consonant with the negative attitudes. This trait is, of course, one that has been commonly observed by writers in this area.

A possibility that is not touched upon by our results is that one of our groups (HL) may represent what Bettelheim and Janowitz call the "stereotyped prejudiced person" (5). These persons might well hold their authoritarian attitude because of an unreflective acceptance of family or community influence. Thus, in the absence of the characterologically based authoritarian needs, a more casual acceptance of our experimental situation could occur. Related to this is the finding reported by Adorno, Frenkel-Brunswick, *et al.* (1) that there is a correlation (particularly in women) between anxiety and authority. Therefore, it is possible that our two authoritarian groups could



be more accurately termed *character-based authoritarians* (with associated anxiety) and *stereotyped authoritarians*.

At the beginning of this study the author planned to run a parallel experiment using the same general procedure, except that there would have been a discussion group that had been instructed to discuss a topic upon which the subject had expressed a strong opinion on a prior questionnaire. The majority would oppose him and the minority member would support him. Although more difficult to control, the more natural sort of activity could lead to interesting material. Unfortunately, some of our students in social psychology became interested in group conformity and, spreading the word about our experiment throughout the campus, destroyed any possibility of carrying it out.

#### F. SUMMARY

The propositions about the behavior of authoritarians who differ with respect to anxiety were investigated in a small group experiment involving a forced alliance between the authoritative subject and the minority figure. A nonauthoritarian group (subdivided as to anxiety) served as a control or reference group.

In this situation, the anxious authoritarians are influenced by majority opinion to a greater degree than are nonanxious authoritarians. Both of the authoritarian groups conform more than do the democratic subjects.

Other propositions relating to attitude changes following the small group experience are partially substantiated. Anxious authoritarians demonstrate a more negative attitude toward the minority person than do an authoritarian control group. However, the predicted generalization of this attitude was only minimally present. In the nonanxious authoritarian group, the expected favorable reaction toward the membership group of the minority participant was not observed. The limited type of interaction involved and the special status of the minority participant are acknowledged as possible explanations for the lack of predicted results.

For the democratic group, the outcome was as predicted on the basis of the results of a prior study on which the procedure in the present study was based.

#### REFERENCES

1. ADORNO, T., FRENKEL-BRUNSWIK, E., LEVISON, D., & SANFORD, N. *The Authoritarian Personality*. New York: Harper, 1950.
2. ALLPORT, G. W. Catharsis and the reduction of prejudice. *J. Soc. Iss.*, 1945, 1 (3), 3-10.
3. ———. *The Nature of Prejudice*. Reading, Mass.: Addison-Wesley, 1945.



4. ASCH, S. *Social Psychology*. New York: Prentice-Hall, 1952.
5. BETTELHEIM, B., & JANOWITZ, M. *Dynamics of Prejudice: A Psychological and Sociological Study*. New York: Harper, 1950.
6. BROPHY, I. The luxury of anti-Negro prejudice. *Publ. Opin. Quart.*, 1946, **10**, 456-466.
7. CARLSON, E. R. Attitude change through modification of attitude structure. *J. Abn. & Soc. Psychol.*, 1956, **52**, 256-261.
8. DROBA, D. Education and Negro attitudes. *Soc. & Sociol. Res.*, 1932, **17**, 137-141.
9. FESTINGER, L. *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row, Peterson, 1957.
10. FESTINGER, L., & KELLY, H. Changing attitudes through social contact. Ann Arbor: Univ. Michigan Research Center for Group Dynamics, 1951.
11. KAGAN, H. E. *Changing the Attitudes of Christian Toward Jew*. New York: Columbia Univ. Press, 1952.
12. MACKENZIE, B. The importance of contact in determining attitudes toward Negroes. *J. Abn. & Soc. Psychol.*, 1948, **43**, 417-441.
13. MCNEMAR, Q. *Psychological Statistics*. New York: Wiley, 1955.
14. MEDNICK, S. Generalization as a function of manifest anxiety and adaptation to psychological experiments. *J. Consult. Psychol.*, 1957, **21**, 491-494.
15. SAMELSON, B. Does education diminish prejudice? *J. Soc. Iss.*, 1945, **1** (3), 11-13.
16. SARASON, I. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403-415.
17. SHERIF, M., & SHERIF, C. *An Outline of Social Psychology*. New York: Harper, 1956.
18. SIEGEL, S. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1956.
19. SMITH, F. T. An experiment in modifying attitudes toward the Negro. No. 837, Contribution to Education, Columbia Univ. Teachers College, New York, 1943. Pp. 80-95.
20. SPENCE, K. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Kentucky Symposium: Learning Theory, Personality Theory, and Clinical Research*. New York: Wiley, 1954.
21. UNITED STATES WAR DEPARTMENT. Opinions about Negro infantry platoons in white companies of seven divisions (Report No. B-157, Army Service Forces). In T. M. Newcomb, E. I. Hartley, et al. (Eds.), *Readings in Social Psychology*. New York: Henry Holt, 1947. Pp. 542-546.
22. WATSON, J. Some social and psychological situations related to change in attitude. *Hum. Relat.*, 1950, **3**, 15-56.
23. YOUNG, D. Some effects of a course in American race problems. *J. Abn. & Soc. Psychol.*, 1927, **22**, 234-242.

*Psychology Department*  
*Woman's Division*  
*Virginia Polytechnic Institute*  
*Radford College*  
*Radford, Virginia*

## LIFE STYLES AND CREATIVITY: A REVIEW\*<sup>1</sup>

*Teachers College, Columbia University; and University of Kansas*

---

BERNARD MACKLER AND FRANKLIN C. SHONTZ

---

### A. INTRODUCTION

A number of studies have utilized life styles as measures of an individual's uniqueness in personality and in perceptual, cognition, and sensory research. Each investigator has tried to grapple with a measure that portrayed uniqueness of the individual yet still differentiated two or more groups of individuals.

Life styles is a concept taken from Adler's notion of "style of life" (3). The self-created style of life is evident in all the functions of the individual; it explains the individuality of the person, his movements, convictions, aspirations, and goals.

Adler described style of life by drawing a comparison with a tree.

If we look at a pine tree growing in a valley we will notice that it grows differently from one on top of a mountain. It is the same kind of a tree, a pine, but there are two distinct styles of life. Its style on top of the mountain is different from its style when growing in the valley. The style of life of a tree is the individuality of a tree expressing itself and molding itself in an environment. We recognize a style when we see it against a background of an environment different from what we expect, for then we realize that every tree has a life pattern and is not merely a mechanical reaction to the environment.

It is much the same with human beings. We see the style of life under certain conditions of environment and it is our task to analyze its exact relation to the existing circumstances, inasmuch as mind changes with alteration of the environment (3, p. 173).

Adler took a strong stand for the idiographic approach by stating that individuals are unique. He was willing to generalize only to the extent of affirming that every person has a style of life and that each life has a unity that pervades all of the individual's thinking, feeling, and acting. Each person molds his life, from constitutional factors and environmental influences, by a

---

\* Received in the Editorial Office on March 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research has been supported in part by funds from the Kansas Neurological Institute. The authors express their gratitude to M. E. Warshauer for her helpful comments and suggestions.

living and a striving for the successful achievement of his self-selected goals. The subjectivity of the individual, his special style, and his approach to life mold and shape all personally relevant events.

The individual, guided by achieving the goal of relating himself to personal problems, and then to social problems, strives to be *superior*. Adler's conception of superiority is personal, unique, subjective; it depends on the meaning the individual gives to his life. Adler felt that every man conducts his life from the beginning as if he knew all his strengths, weaknesses, abilities, and disabilities. A person's behavior springs from his opinion of what needs to be done and how he can do it.

The development of the style of life begins in childhood. In that formative stage, the body exerts the greatest influence, for the sense and bodily organs are felt by the individual in his movement and functions. The child, by trial and error, explores his body and his environment. He feels his limitations and strives to overcome his shortcomings, and his purpose in life becomes one of attaining mastery over himself—to become a whole being. The particular meanings he comes to assign to his life on the basis of his strivings and his experiences determine how he will strive to complete himself through the rest of his personal existence.

Wertheimer once formulated a principle that bears some similarity to Adler's style-of-life concept. According to an unpublished statement quoted by Allport, Wertheimer held that "if a personality is rightly understood, it will always be found to lie under the domination of one controlling goal, one ruling passion, possessing one *radix* . . ." (3, p. 175).

Although Adler stressed the uniqueness of individual styles of life, it is not unreasonable to suppose that such styles may be examined objectively and classified into relatively homogeneous groups. Indeed, there are trends in behavioral science suggesting that such developments are already taking place. Mackler (20) has completed a study in which he assessed five "life-style" groups of undergraduate students. He compared art majors, dance majors, physically disabled, visually disabled students, and a control group on measures of sensory life style and creativity. Life styles posed a measurement problem in that a person's stated major or vocational interest appeared unreliable as a measure of life style. A scale was developed, the Life Style Scale, to tap specifically visual and kinesthetic life styles. The scale lived up to expectations, for it readily discriminated the most visual group (art majors) from the rest and also the most kinesthetic group (dance majors) from the rest.

Several investigators interested in perceptual and cognitive differences among subjects have postulated concepts reminiscent of Adler's style of

life. For example, Witkin (35) has studied differences in modes of perception and has described two broad perceptual styles: field-dependent and field-independent. Those individuals who, on a perceptual task—the rod-and-frame test—rely heavily on the surrounding field and are unable to overcome it are considered field dependent. Contrasted with this performance, field-independent subjects are able to bring the rod close to the true vertical position without being heavily influenced by the surrounds, but rather by personal body position. Witkin and his colleagues (36) document how these perceptual styles exist for these two broad types and also how other personality, intellectual, and social measures of behavior differentiate the two groups. The consistency of differences pervades the functions to the extent that the field-dependent individual is easily influenced by authority figures, is attentive to the faces of those around him, favors professions that deal with people, is popular with the group, and has difficulty in locating a familiar figure in a complex design. In their behavior characteristics, field-independent individuals show contrasts that are not always exactly opposite to those of field-dependent individuals. Compared to field dependents, however, field independents are more autonomous; some are loners, overcontrolled, removed, cold, and distant. Witkin *et al.* (36, p. 4) conclude:

With findings such as these, an investigation that had dealt with individual differences in a seemingly narrow perceptual activity developed into a study of broad differences among people in what seemed to add up to a "style of life."

In cognition research, Klein (15) postulated "cognitive style" to account for the self-consistency in a person's behavior. He defined cognitive style as the "structural arrangement of cognitive attitudes." Cognitive attitudes are involved in adaptation to the environment, while cognitive styles encompass behavior and:

... reflect a superordinate level of control within the organism. One basis for conceiving of this possibility is to think of certain types of behavior as a "secondary adaptation" to a cognitive attitude, after continual and repeated reliance upon the attitude (15, p. 11).

Gardner (5) with his associates (6, 7), investigating cognitive controls, has been heavily influenced by Klein's principle of cognitive style. For Gardner, style answers the issue of typology in explaining differences in cognition. Klein (15) and Gardner *et al.* (6) have suggested that molar aspects of individual differences in other areas of personality organization are linked primarily to styles.



Getzels and Jackson (9) have also used this concept of cognitive style in their research. They were concerned with the family environment and its effect on the cognitive style of children. The families of high-*IQ* and high-creativity adolescents were studied intensively to discover the possible relationships of family environment on two types of cognitive styles: i.e., intelligence and creativity.

### B. CREATIVITY AND LIFE STYLES

The theorists previously discussed also introduced creativity as a factor in the development of the style of life. Adler was the foremost in developing an interrelationship between creativity and life style. At times, Adler (3) did not differentiate his concepts of "the creative self" and "style of life," for he saw these as interwoven; creativity provides meaning and substance to life. Creativeness allows man to *make* and to develop his own personality, his approach to life. This stylization or way of life is created, molded, carved out, and individualized by each person. This style of life is mirrored in an individual's creative endeavors, be they maximal or minimal, in building his unique personality. As Adler said:

... each individual case will be found to have developed its own unique style of life, the origin of which is ultimately to be sought in the individual's own creativeness (3, p. 240).

The individual, as Adler postulates, is always striving to complete himself. His emergent life is always driven by a sense of incompleteness and he tries, desperately at times, to forge ahead to *create* a whole being—to actualize himself.

Mackler (20) utilized Adler's concepts of life style and creativity in a study devoted to a comparison of the interaction of an individual's life style and his creative expression. He compared undergraduate students who were art majors, dance majors, physically disabled, visually disabled, and a control group that excluded art majors, dance majors, and disabled students. The results are noteworthy in that they point to how poorly the art group is discriminated from the remaining groups. Aside from gross combinations of 14 creativity scores, only one variable—the *Circles* originality score—proved to be an effective discriminator with respect to the art students versus the four remaining groups. The most illuminating finding was the performance of the physically disabled, significantly the lowest on each of seven variables, and the meaning this may have in relating severe disability to creativeness. This result points to factors in the life style of the individual that, in the



physically disabled, tend to thwart creativity and tend to cause the choice of conformity as a solution if not as a desired end.

Adler described the interaction of creativity with life style in a social, interpersonal context. Getzels and Jackson (9), in their study of family environment and its effect on creativity and intelligence, found that there is familial, social influence that affects the cognitive style of the children and hence creativity. Families of high-*IQ* students seem to be more "vigilant" than others with respect to their children's behavior and their manifest academic performance. Also, they tend to be critical of their children, the school, and their children's friends. In contrast, the parents of high-creative adolescents focus on less visible qualities, such as their child's openness to experience, his values, his interests, and his enthusiasms.

Torrance (33) studied cross-cultural differences in creativity. The data were obtained from a large school area in Midwestern United States, from Australia, Western Samoa, Germany, India, and from segregated Negro schools in Georgia. The children were tested with three nonverbal tasks developed by Torrance: Picture Construction, Incomplete Figure, and Circles and Squares. In each of the cultures studied, approximately 1,000 pupils in grades one through six were examined. Developmental curves for each culture were compared and the results were seen as supporting Mead's views (21) that cultures affect creativity differently. The patterns of development appeared similar for the United States and India, although Indian children scored lower at all grade levels. Samoan children showed a consistent increase from grade level to grade level, although they scored much lower than the remaining groups. All the groups, other than those from Samoa, showed vast drops in creativity at certain grade levels, especially at the fourth grade.

Anderson (1), incorporating the views of Fromm, Rogers, and Maslow, described a developmental process. He saw the family and social environment affecting the child's personality from birth on. He described the growth pattern and also the effects of child-rearing practices on creativity. In his developmental theory, Anderson asked the question: what happens to children born with qualities and capacities for creativity when they become adults? Adults are not usually creative; for the most part, they are passive and fearful conformists. He postulated that children, in the process of development, lose their creative urges by conforming; their life aspirations change and so do their initiating propensities.

John Anderson (2), in agreement with Harold Anderson (1), stated that creativity is a result of a long-term functioning-process. Complex forms of creativity do not involve merely a momentary inspiration, but are concerned

with the organization of ideas, the structuring of concepts, and the selection of relations over time. Cox (4), Roe (26), and Stein *et al.* (28) have completed studies that do justice to both Andersons in terms of studying the life histories of talented scientists. They aimed at understanding the developmental life process and its effect on creativity. They studied talented individuals biographically.

Personality tests have also been used to study creative professional persons. MacKinnon (19) studied architects by means of group tests: the Minnesota Multiphasic Personality Inventory (MMPI), the California Psychological Inventory, the Allport-Vernon-Lindzey Study of Values, and the Strong Vocational Interest Blank. MacKinnon described the architects in terms of personality and vocational patterns evinced by these tests. Other studies using psychological tests to identify creative individuals were undertaken by Hammer (13) and by Taylor (29).

Two studies, one by Hebeisen (14) and one by Tisdall (30), used tests developed by Torrance to investigate differences in creativity with respect to the emotionally disturbed and mentally retarded as compared to "normals." Hebeisen (14) studied schizophrenics; Tisdall (30) studied the mentally retarded. Hebeisen found that the disturbed group showed a marked impoverishment of imaginative and of creative thinking abilities. Tisdall found that educable mentally retarded children in special classes not only performed on tests of creativity, but they did as well as intellectually normal subjects. This result reiterates the finding that intelligence is not highly correlated to creativity, as Torrance (34), Yamamoto (38), and Guilford (12) have previously pointed out.

### C. CREATIVITY AND SENSORY LIFE STYLES

Schachtel (27) has discussed the impact of sensory stimuli and its effect on the individual. Murphy (22) was somewhat more specific than Schachtel in his discussion of "sensitiveness." For Murphy, sensory stimulation and enrichment are the essential experiences that can give rise to creation:

The first great phase in the evolution of the creator appears to be the extreme *sensitiveness* to a specific form or experience, usually sensory; it is especially likely to involve sight, sound, or the muscle sense. It embodies delight in these experiences, a need for more, a curiosity into their relationships; in other words, we are dealing with sensory and activity drives. Such a person dwells more and more upon these experiences, exposes himself to them, and as far as possible controls them so that at will he can have what he wants. He tries them in combination and selects the most delightful combination; creation thus flows out of the original need for a form of experience (22, p. 452).

Sensory experiences, important as they are, cannot by themselves create. Murphy added that for creativity to occur there must be a discovery of relations between sensory experiences and a manipulation of these experiences so as to restructure their relations, to complicate them, and to improvise new ones.

Lowenfeld (17, 18) extended Murphy's ideas about sensory experiences to the point where he differentiated two types of sensory perceivers, hence two sensory styles. He stated that there are "visual" and "haptic" ways of perceiving. The two different kinds of perception correspond to two different creative types: the visual type is mainly concerned with his environment; the haptic type, with his own bodily sensations and with the tactual space about him. The visually minded person depicts what he sees; he is affected by what is out there, the external, and he tries to represent the environment analytically. The nonvisually minded person reflects his emotions, his feelings, his body; he depicts what he feels inside and his creation reflects his emotional and bodily reactions.

The term "visual perception" is used when visual impressions are the dominant features in a percept and they subordinate the impressions coming from the other senses. "Haptic perception" is the synthesis between tactile perception and external reality and those subjective experiences that are closely bound with the experience of self, such as bodily sensation, muscular innervations, deep sensibilities, and their emotional affect. These two styles, visual and haptic, are not pure types because impressions coming from a given sense are always influenced more or less by impressions coming from other senses. However, they do represent extreme differences resulting from psychological factors and not from physiological factors.

Lowenfeld (18) and Revesz (25) also discussed visual and haptic perception with respect to creativity in the partially and the totally blind. For the partially blind (18), both visual and haptic types exist, for the visual type can still use his sight. For the totally blind (25), there is almost total reliance on feelings, and these blind create out of their bodily sensations. The blind use their kinesthetic and tactile impressions and are therefore haptic.

Sensory factors—be they visual, haptic, auditory, tactile, or olfactory—play an important part in creativity. Myers and Torrance (23), in their workbook designed to foster creativity in elementary-school children, include a unit devoted exclusively to sense impressions. The unit, called "Take A Sensible Trip," asks the children to think and to explore their sense impressions on an imagined trip to home. The students are encouraged to use all their senses:

Use your nose, your hands, your ears, and your eyes—make use of *all* your senses. See if there are differences in the atmosphere, in the movement of things, in your attitude towards the world, and in its attitude towards you (23, p. 5).

A few empirical studies have been undertaken to test the effect of sensation upon creativity. Littwin (16) tried three methods of developing imaginative writing with matched groups of seventh and eighth graders. During a period of 10 weeks, the first group practiced writing vivid descriptions. The second group studied literary models containing words of sound, color, and movement. The third group practiced describing all the possible sensations (such as sight, smell, touch, and hearing) that they could experience in examining an object or situation. Those in the latter group showed significantly greater gains on composition tests than did the other two groups.

Witty (37) used films in an attempt to stimulate elementary-grade pupils visually. He used a film entitled "The Hunter and the Forest—A Story Without Words" to elicit reactions in terms of sensitivity in expression, expression of genuine feeling, and appreciation for the intent of the filmmaker. Of 2,000 compositions written by these grade-school pupils (second to sixth grades) following the showing of the film, Witty judged that 10 per cent showed superior promise in writing. The study was not controlled, for the students saw the film but once and the only evaluation of their writing was made after the film presentation.

Both of these studies by Littwin and Witty focused on sensory stimulation and its effect on creativity. Goldberger and Holt (11) studied the effects of sensory deprivation. They found that the subjects who, on the Rorschach, handled primary process material in a mature and effective way reacted adaptively to isolation. The subjects who (on the Rorschach) handled primary process material with poor control reacted negatively to isolation. The former group had pleasant effects, evinced rational thinking, and perceived the situation as not threatening, while the latter group had unpleasant effects, anxiety-laden intrusions of primary process, and difficulty in rational thinking.

#### D. SUMMARY

Adler's concept of life style has been discussed in detail. Its utilization as a sound concept is evident not only for Adlerian clinicians, but also for empirical investigators.

The second section of the paper reviewed Adler's contribution to the interrelationship between creativity and life style. Contemporary research



related to Adler's formulation was presented to document the efficacy of the conceptual relationship between creativity and life style.

The last section of the paper was devoted to one specific component of life style, the sensory mode. Sensory life style was discussed with specific relationship to creative expression.

#### REFERENCES

1. ANDERSON, H. H., *Ed.* Creativity and Its Cultivation. New York: Harper, 1959.
2. ANDERSON, J. E. The nature of abilities. In E. P. Torrance (Ed.), *Education and Talent*. Minneapolis: Univ. Minnesota Press, 1960. Pp. 9-31.
3. ANSBACHER, H. L., & ANSBACHER, R. R., *Eds.* The Individual Psychology of Alfred Adler. New York: Basic Books, 1956.
4. COX, C. M. Crucial factors in the life history of talent. In E. P. Torrance (Ed.), *Education and Talent*. Minneapolis: Univ. Minnesota Press, 1960. Pp. 51-65.
5. GARDNER, R. W. Cognitive control principles and perceptual behavior. *Bull. Menninger Clin.*, 1959, **23**, 241-248.
6. GARDNER, R. W., JACKSON, D. N., & MESSICK, S. J. Personality organization in cognitive control and intellectual abilities. *Psychol. Iss.*, 1960, **2**, No. 8.
7. GARDNER, R. W., HOLTZMAN, P. S., KLEIN, G. S., LINTON, H., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Iss.*, 1959, **1**, No. 4.
8. GETZELS, J. W., & JACKSON, P. W. The social context of giftedness: A multi-dimensional approach to definition and methods. In E. P. Torrance (Ed.), *Creativity: Second Minnesota Conference on Gifted Children*. Minneapolis: Center for Continuation Study, Univ. Minnesota, 1959. Pp. 28-34.
9. ———. Family environment and cognitive style: A study of sources of highly intelligent and highly creative adolescents. *Amer. Sociol. Rev.*, 1961, **26**, 351-359.
10. ———. Creativity and Intelligence: Explorations with Gifted Students. New York: Wiley, 1962.
11. GOLDBERGER, L., & HOLT, R. R. Experimental interference with reality contact: Individual differences. In P. Solmon, P. E. Kubzansky, P. H. Liederman, J. H. Mendelson, R. Trumbull, & D. Wexler (Eds.), *Sensory Deprivation*. Cambridge, Mass.: Harvard Univ. Press, 1961. Pp. 130-142.
12. GUILFORD, J. P. Traits of creativity. In H. H. Anderson (Ed.), *Creativity and Its Cultivation*. New York: Harper, 1959. Pp. 142-161.
13. HAMMER, E. F. Creativity. New York: Random House, 1961.
14. HEBEISEN, A. A. The performance of a group of schizophrenic patients on a test of creative thinking. In E. P. Torrance (Ed.), *Creativity: Second Minnesota Conference on Gifted Children*. Minneapolis: Center for Continuation Study, Univ. Minnesota, 1959. Pp. 125-129.
15. KLEIN, G. S. Cognitive control and motivation. In G. Lindzey (Ed.), *Assessment of Human Motives*. New York: Holt, Rinehart, & Winston, 1961. Pp. 87-118.
16. LITWIN, M. An experimental investigation of the effect of method of presentation upon the imaginative quality of descriptive writing among elementary-school pupils. *J. Exper. Educ.*, 1935, **4**, 44-49.
17. LOWENFELD, V. The Nature of Creative Activity. London: Kegan Paul, Trench, & Trubner, 1939.
18. ———. Creative and Mental Growth. A Textbook on Art Education. New York: Macmillan, 1947.



19. MACKINNON, D. W. The nature and nurture of creative talent. *Amer. Psychologist*, 1962, 17, 484-495.
20. MACKLER, B. Creativity and life styles. Unpublished Doctoral dissertation, University of Kansas, Lawrence, 1963.
21. MEAD, M. Creativity in cross-cultural perspective. In H. H. Anderson (Ed.), *Creativity and Its Cultivation*. New York: Harper, 1959. Pp. 222-235.
22. MURPHY, G. Personality: A Biosocial Approach to Origins and Structure. New York: Harper, 1947.
23. MYERS, R. E., & TORRANCE, E. P. Invitation to Thinking and Doing. Minneapolis: Perceptive Publishing, 1961.
24. ———. Teachers' Manual for Invitations to Thinking and Doing. Minneapolis: Perceptive Publishing, 1961.
25. REVESZ, G. Exploring the World of Touch: Psychology and Art of the Blind. London: Longmans & Green, 1950.
26. ROE, A. Crucial life experiences in the development of scientists. In E. P. Torrance (Ed.), *Talent and Education*. Minneapolis: Univ. Minnesota Press, 1960. Pp. 66-77.
27. SCHACHTEL, E. G. Metamorphosis. New York: Basic Books, 1959.
28. STEIN, M. I., MACKENZIE, J. N., RODGERS, R. R., & MEER, B. A case study of a scientist. In A. Burton & R. E. Harris (Eds.), *Clinical Studies of Personality*. New York: Harper, 1955. Pp. 726-767.
29. TAYLOR, C. W. A tentative description of the creative individual. In W. B. Waetjen (Ed.), *Human Variability and Learning*. Washington: National Education Assoc., 1961. Pp. 62-79.
30. TISDALL, W. J. Productive thinking in retarded children. *Exceptional Child*, 1962, 29, 36-41.
31. TORRANCE, E. P., Ed. Creativity: Second Minnesota Conference on Gifted Children. Minneapolis: Center for Continuation Study, University of Minnesota, 1959.
32. ———. Personality development of the highly creative child. In E. P. Torrance (Ed.), *Creativity: Second Minnesota Conference on Gifted Children*. Minneapolis: Center for Continuation Study, University of Minnesota, 1959. Pp. 98-100.
33. ———. Cultural discontinuities and the development of originality of thinking. *Exceptional Child*, 1962, 29, 2-13.
34. ———. Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice-Hall, 1962.
35. WITKIN, H. A. The nature and importance of individual differences in perception. *J. Personal.*, 1949, 18, 145-170.
36. WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. Psychological Differentiation. New York: Wiley, 1962.
37. WITTY, P. The use of films in stimulating creative expression and in identifying talented pupils. *Elementary English*, 1956, 33, 340-344.
38. YAMAMOTO, K. Creativity and intellect: A review of current research and projection. Paper read at Minnesota Psychological Association, Minneapolis, April, 1961.

*Institute of Urban Studies  
Teachers College  
Columbia University  
New York 27, New York*

HYPNOTIZABILITY, SUGGESTIBILITY, AND PERSONALITY:  
I. TWO STUDIES WITH THE EDWARDS PERSONAL PREFERENCE  
SCHEDULE, THE JOURARD SELF-DISCLOSURE SCALE,  
AND THE MARLOWE-CROWNE SOCIAL DESIRABILITY  
SCALE\* <sup>1</sup>

*Medfield Foundation, Harding, Massachusetts*

---

THEODORE XENOPHON BARBER AND DAVID SMITH CALVERLEY

---

A. INTRODUCTION

This paper presents two studies designed to test the hypothesis that individual differences in hypnotizability and suggestibility are related to differences among individuals in personality characteristics. Three personality instruments were used: the Edwards Personal Preference Schedule (16), the Jourard Self-Disclosure Scale (18), and the Marlowe-Crowne Social Desirability Scale (15).

The Edwards Personal Preference Schedule (EPPS) purports to measure 15 "manifest personality needs," such as Dominance, Autonomy, Affiliation, and Deference. The inventory was constructed in a manner designed to minimize tendencies for Ss to claim desirable characteristics and to deny undesirable ones. Using the postural-sway test to measure suggestibility in 27 student nurses, Zuckerman and Grosz (24) found that the suggestible obtained significantly lower scores than the nonsuggestible on the Autonomy scale of the EPPS.

Lang and Lazovik (19)—and Levitt, Brady, and Lubin (20)—administered the EPPS to 32 college students and to 31 nursing students respectively; then both groups of Ss were classified on response to the standardized-test suggestions included in the Stanford Hypnotic Susceptibility Scale. Lang and Lazovik reported a significant correlation (.36) between hypnotizability and

---

\* Received in the Editorial Office on April 6, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by grants MY-04825 and MH-07003 from the National Institute of Mental Health, United States Public Health Service.

We wish to express our appreciation to the following for their cooperation in obtaining subjects: Emily Ross, Dean of Women at Becker Junior College; Professor Howard H. Chauncey of Tufts School of Dental Medicine; Harry Y. Hilyard, Superintendent, Waldren P. Lojko, Principal, and Wallace Johnson, Vice Principal, of Grafton (Mass.) Junior-Senior High School; Doris Heusel, Supervisor of Student Nurses at Medfield State Hospital; and John Elberfeld, Dean, and Kenneth Wolkon, Instructor, at Worcester Junior College.

Affiliation, and Levitt *et al.* reported that the hypnotizable show greater "dependency" than the nonhypnotizable—as indicated by lower scores on the combined Autonomy, Dominance, and Aggression scales. In administering the EPPS we expected to find that the hypnotizable and suggestible would obtain higher scores than the nonhypnotizable and nonsuggestible on the Affiliation scale and lower scores on the Autonomy, Dominance, and Aggression scales.

The Jourard Self-Disclosure Scale is designed to determine to what extent the *S* has confided or disclosed personal information about himself to four specified individuals: his mother, father, closest friend of the same sex, and closest friend of the opposite sex. The questionnaire lists items of personal information concerning attitudes and opinions, tastes and interests, feelings about work and money affairs, and worries or concerns about self and physical appearance. The *S* responds to each item by indicating whether he has disclosed the personal information to each of the four specified individuals. Jourard<sup>2</sup> had advanced the hypothesis that hypnotizable individuals are more willing than the nonhypnotizable to confide or to disclose themselves to others. This hypothesis was put to empirical test in the present investigation.

The Marlowe-Crowne Scale is designed to measure "social desirability" (defined as a need to obtain approval from others by responding in a socially appropriate and acceptable manner). The inventory contains 33 items which refer to behaviors that are culturally approved and socially desirable and yet which cannot be endorsed by most people if they are answering truthfully (e.g., "No matter who I'm talking to, I'm always a good listener"). We administered the Marlowe-Crowne inventory under the implicit hypothesis that need to obtain approval from others is positively related to hypnotizability and suggestibility.

## B. STUDY I: HYPNOTIZABILITY AND PERSONALITY

### 1. Method

Ss participating in Study I were 100 undergraduates (57 males and 43 females) from two colleges. The males were required to participate in the experiment to fulfill the requirements of an elementary psychology course. The females were volunteers (paid one dollar).

All Ss first completed the EPPS. In a second session, each *S* was rated individually on *hypnotizability*: a standardized 15-minute hypnotic-induction procedure was administered, and response to the Barber Suggestibility Scale

<sup>2</sup> Jourard, S. M. Personal communication to T. X. Barber, September 17, 1960.

(6, 13, 14) was then assessed.<sup>3</sup> The hypnotic-induction procedure was patterned after the hypnosis procedures of Friedlander and Sarbin (17), Marcuse (21), and Weitzenhoffer and Hilgard (23). In a third session all Ss completed the Jourard Scale; 83 Ss also completed the Marlowe-Crowne Scale.

Pearsonian correlations were computed (separately for each sex) between Objective scores on the Barber Suggestibility Scale and scores on each of the following: the Consistency scale and the 15 personality dimensions of the EPPS; the five variables measured by the Jourard Scale (self-disclosure to mother, father, same-sex friend, opposite-sex friend, and scores on the total scale); and total scores on the Marlowe-Crowne Scale.

## 2. Results

No significant relationships were found between hypnotizability and any of the variables measured by the EPPS, the Jourard Scale, or the Marlowe-Crowne Scale. Of the 22 correlations computed for the males, none accounted for more than 6 per cent of the variance and none differed significantly from zero. Of the 22 correlations computed for the females, none accounted for more than 7 per cent of the variance and none differed significantly from zero.

## C. STUDY II: SUGGESTIBILITY AND PERSONALITY

### 1. Method

Ss participating in Study II were 414 students (262 males and 152 females) of whom 205 were college undergraduates or dental-school post-graduates, 95 were nursing students, and 114 were high-school students. As Table 1 (Row 5) indicates, 149 Ss were volunteers (paid one dollar) and 265 were required to serve as subjects either to fulfill the requirements of a course in elementary psychology (male college students) or to comply with the direct request of the school principal and the study-hall teacher (high-school students). The investigation was defined to all Ss as a test of imagination and personality.

In the first session, the Ss completed the EPPS. In the second session, they were rated individually on *suggestibility* after administration of the Barber Suggestibility Scale *without* a preceding hypnotic-induction procedure. Immediately prior to assessment of suggestibility, 124 Ss were given brief Task-

<sup>3</sup> All Ss participating in Study I and also all Ss participating in Study II were tested by one experimenter (DSC).



Motivating instructions, as described elsewhere (4, 7), and the remaining 290 were *not* given such instructions (Direct-Suggestions condition). In a third session, 300 Ss completed the Jourard Scale and 137 also completed the Marlowe-Crowne Scale.<sup>4</sup>

Pearsonian correlations between Objective scores on the Barber Suggestibility Scale and scores on each of the variables measured by the three personality inventories were computed separately for each sex group tested under the Task-Motivation and the Direct-Suggestions conditions. All Ss completing the inventories were included in these computations with the exception of those who were inconsistent in their responses to the EPPS (scoring 8 or below on the Consistency scale of this inventory); these Ss were not included in the analyses of the 15 EPPS variables.

## 2. Results

Table 1 shows that nine correlations between suggestibility and the EPPS variables are significantly different from zero at the .05 level. These results barely exceed chance expectations (as 112 correlations were computed with respect to the EPPS, about six should have been significant at the .05 level by chance alone). The significant correlations do not show a consistent pattern. Of the seven groups assessed in the study, two groups show significant negative correlations between suggestibility and need for Achievement; one group shows a significant negative correlation between suggestibility and need for Change; and one group shows a significant positive correlation between suggestibility and scores on the Consistency scale. Three correlations between suggestibility and Autonomy are significant; two of these are negative and one is positive. Two correlations between suggestibility and Nurturance are significant; one of these is negative and one is positive.

Table 1 shows that one correlation between suggestibility and the Jourard Self-Disclosure Scale is significant at the .05 level. (Because 25 correlations were computed with respect to this scale, at least one should have been significant by chance alone.)

Table 1 also shows that none of the correlations between suggestibility and scores on the Marlowe-Crowne Social Desirability Scale are significantly different from zero.

## D. DISCUSSION

Little if anything can be concluded from these negative findings taken by

<sup>4</sup> Some of the Ss participating in Studies I and II also completed the As (1), the Shor-Orne-O'Connell (22), and the Barber-Glass (13) questionnaires. The negative results obtained with these questionnaires are presented and discussed elsewhere (10).



TABLE 1  
CORRELATIONS BETWEEN SUGGESTIBILITY AND SCORES ON THE EDWARDS PERSONAL PREFERENCE SCHEDULE, THE JOURNAL SELF-DISCLOSURE SCALE, AND THE MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE

Variable	Experimental condition			
	Task Motivation "Captive" male college students	"Captive" male high-school students	"Captive" female high-school students	Direct Suggestions Volunteer nursing students
Consistency	(N = 63) .37*	(N = 61) .20	(N = 57) -.10	(N = 34) .09
15 EPPS variables	(N = 58)	(N = 58)	(N = 50)	(N = 34)
Achievement	-.13	.06	-.21	-.40*
Deference	-.09	.23	.06	.00
Order	.02	.17	-.03	-.06
Exhibition	.23	.07	.18	-.11
Autonomy	.03	-.31*	-.13	-.20
Affiliation	-.13	.04	.16	.29
Intracception	.24	.06	.02	.13
Succorance	-.01	.01	.05	.17
Dominance	.05	-.20	.07	-.24
Abasement	.04	.09	.17	.30
Nurturance	.09	-.10	.19	.37*
Change	-.26	-.01	-.32*	.09
Endurance	-.12	.17	-.24	-.14
Heterosexuality	-.07	-.03	.00	-.03
Aggression	-.06	-.14	-.02	-.20
Self-Disclosure variables	(N = 63)	(N = 61)	<i>Journal Self-Disclosure Scale</i>	
Mother	-.02	-.19	(N = 34)	(N = 34)
Father	.09	-.11	-.13	-.19
Same-sex friend	-.20	-.17	-.09	.26
Opposite-sex friend	-.14	-.15	-.10	.07
Total	-.02	-.33*	-.24	-.01
			-.22	.11
Social Desirability variable	(N = 28)	(N = 31)	<i>Marlowe-Crowne Social Desirability Scale</i>	
Total	.21	.15	(N = 27)	(N = 27)
			-.36	.03

\*  $p < .05$ .

themselves. Further, little if anything can be concluded from a series of recent investigations (2, 9, 10, 11) in which we failed to find reliable relations between hypnotizability or suggestibility and other personality instruments, such as the Leary Interpersonal Check List, the Rosenzweig Picture-Frustration Study, and the questionnaires constructed by As (1) and by Shor, Orne, and O'Connell (22). However, our consistently negative findings, *taken in the context of previous research*, appear to indicate an important conclusion relevant to theories of suggestibility and hypnotic behavior. Previous research, reviewed in detail elsewhere (3), either failed to exhibit reliable relations between personality characteristics and hypnotizability or suggestibility, or presented conflicting "significant" findings. These negative and contradictory results (obtained in an extensive series of studies conducted by a large number of investigators) suggest that personality characteristics—as measured by self-report inventories, projective tests, clinical interviews, and other present-day methods of assessment—may not by themselves play a major role in hypnotizability or suggestibility.

If personality characteristics *per se* do not account for the major portion of the variance, what factors are of major importance? Recent experiments (4, 5, 7, 8, 12) indicate that interindividual differences and intraindividual consistency in response to suggestions may be due primarily to factors that are more situationally variable than those traditionally subsumed under the concept of personality characteristics. These factors appear to include (a) *S*'s attitude toward and relationship with the experimenter, (b) his attitudes and goals in the immediate experimental situation, and (c) his motivation to perform well or poorly on assigned tasks. We would suggest that further research attempting to assess the relevance of personality characteristics to hypnotizability and suggestibility should focus on the *interactions* between *Ss*' personality characteristics and situational variables, such as the following (each of which appears to affect *Ss*' attitudes and motives toward the immediate suggestive situation): what the *S* is told about the experiment, the characteristics of the experimenter as perceived by the *S*, and the interpersonal interaction between the *S* and the experimenter that occurs prior to assessment of suggestibility.

#### E. SUMMARY

Two studies, with 100 and 414 *Ss* respectively, failed to demonstrate that hypnotizability and suggestibility are related to need to obtain approval from others (Marlowe-Crowne Social Desirability Scale), to willingness to confide to others (Jourard Self-Disclosure Scale), or to Deference, Autonomy, Affili-

ation, Dominance, Abasement, Nurturance, Aggression, or any of the other personality variables measured by the Edwards Personal Preference Schedule. The results are related to previous work in this area which also yielded negative or contradictory findings. It is suggested that further research should attempt to specify the effects on hypnotizability and suggestibility of personality variables *in interaction with* situational variables.

## REFERENCES

1. AS, A. Non-hypnotic experiences related to hypnotizability in male and female college students. *Scand. J. Psychol.*, 1962, **3**, 112-121.
2. BARBER, T. X. Failure of the Rosenzweig Picture-Frustration Test to Relate to Primary Suggestibility. Harding, Mass.: Medfield Foundation, 1961.
3. ———. Hypnotizability, suggestibility, and personality: V. A critical review of research findings. *Psychol. Rep.*, 1964, **14**, 299-320.
4. BARBER, T. X., & CALVERLEY, D. S. "Hypnotic behavior" as a function of task motivation. *J. of Psychol.*, 1962, **54**, 363-389.
5. ———. Experimenter's Tone of Voice as a Variable Affecting "Hypnotic-Like" Suggestibility. Harding, Mass.: Medfield Foundation, 1963.
6. ———. "Hypnotic-like" suggestibility in children and adults. *J. Abn. & Soc. Psychol.*, 1963, **66**, 589-597.
7. ———. The relative effectiveness of task-motivating instructions and trance-induction procedure in the production of "hypnotic-like" behaviors. *J. Nerv. & Ment. Dis.*, 1963, **137**, 107-116.
8. ———. Toward a theory of hypnotic behavior: Effects on suggestibility of task motivating instructions and attitudes toward hypnosis. *J. Abn. & Soc. Psychol.*, 1963, **67**, 557-565.
9. ———. Hypnotizability, suggestibility, and personality: III. A study using teachers' ratings of children's characteristics. *J. of Psychol.*, 1964, **57**, 275-280.
10. ———. Hypnotizability, suggestibility, and personality: II. Assessment of previous imaginative-fantasy experiences by the As, Barber-Glass, and Shor questionnaires. *J. Clin. Psychol.*, in press.
11. ———. Hypnotizability, suggestibility, and personality: IV. A study with the Leary Interpersonal Check List. *Brit. J. Soc. & Clin. Psychol.*, in press.
12. ———. Toward a theory of "hypnotic" behavior: Effects on suggestibility of defining the situation as hypnosis and defining response to suggestions as easy. *J. Abn. & Soc. Psychol.*, in press.
13. BARBER, T. X., & GLASS, L. B. Significant factors in hypnotic behavior. *J. Abn. & Soc. Psychol.*, 1962, **64**, 222-228.
14. BARBER, T. X., KARACAN, I., & CALVERLEY, D. S. "Hypnotizability" and suggestibility in chronic schizophrenics. *Arch. Gen. Psychiat.*, in press.
15. CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. Consult. Psychol.*, 1960, **24**, 349-354.
16. EDWARDS, A. L. Revised Manual for the Edwards Personal Preference Schedule. New York: Psychological Corp., 1959.
17. FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *J. Abn. & Soc. Psychol.*, 1939, **33**, 453-475.
18. JOURARD, S. M., & LASAKOW, P. Some factors in self-disclosure. *J. Abn. & Soc. Psychol.*, 1958, **56**, 91-98.
19. LANG, P. J., & LAZOVIK, A. D. Personality and hypnotic susceptibility. *J. Consult. Psychol.*, 1962, **26**, 317-322.

20. LEVITT, E. E., BRADY, J. P., & LUBIN, B. Correlates of hypnotizability in young women: Anxiety and dependency. *J. Personal.*, 1963, **31**, 52-57.
21. MARCUSE, F. L. Hypnosis: Fact and Fiction. Baltimore: Penguin Books, 1959. Pp. 52-53.
22. SHOR, R. E., ORNE, M. T., & O'CONNELL, D. N. Validation and cross-validation of a scale of self-reported personal experiences which predicts hypnotizability. *J. of Psychol.*, 1962, **53**, 55-75.
23. WEITZENHOFFER, A. M., & HILGARD, E. R. Stanford Hypnotic Susceptibility Scale. Palo Alto, Calif.: Consulting Psychologists Press, 1959. Pp. 13-18.
24. ZUCKERMAN, M., & GROSZ, H. J. Suggestibility and dependency. *J. Consult. Psychol.*, 1958, **23**, 328.

*The Medfield Foundation, Inc.*  
*Harding, Massachusetts*

## VIGILANCE IN COMPLEX TASK SITUATIONS\*<sup>1</sup>

*U. S. Army Medical Research and Development Command, Lockheed-Georgia Company, and University of Louisville*

GLENN R. HAWKES, THOMAS W. MEIGHAN, AND EARL A. ALLUISI

### A. INTRODUCTION

Investigations of man's monitoring abilities typically have demonstrated that performance efficiency decreases as a function of time-on-task (12). With few exceptions, such investigations have stressed efficiency in terms of the detection of some sort of critical signal under relatively monotonous conditions of performance. Some studies have used more than one kind of sensory input in a single session; for example, visual and auditory signals have been used (4, 5, 6, 13), as have auditory and cutaneous stimuli (11). Under conditions permitting more than one kind of stimulus to be presented within the session, the resultant probability of signal detection may or may not be improved over that obtained when only one kind of signal is used. There have been no systematic attempts to vary the level of performance demanded of the subjects in order to determine directly the effect of task demand on monitoring efficiency.

A task program of the sort to which reference has just been made should ideally encompass (a) several levels of performance demand so arranged as to permit progress from one level to another and a return to the original level at least once during a given session, (b) several different kinds of tasks that—collectively—sample as many different abilities as possible, and (c) tasks on which performance data indicate an acceptable degree of measurement reliability. The program shown in Table 1 has been used in several studies to assess individual and group abilities for periods of as long as 30 days of sustained performance (3).

The tasks listed in Table 1 have been described elsewhere (3). The three tasks presented during only part of a session may be characterized as mental arithmetic, the processing of sequentially encoded information, and form perception. In addition, three monitoring tasks were presented continuously.

\* Received in the Editorial Office on April 13, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Based on data recorded at the Human Factors Research Laboratory, Lockheed-Georgia Company, Marietta, Georgia. A paper based on these data was read to the Psychonomic Society, Bryn Mawr College, Pennsylvania, August 31, 1963.





All tasks were presented to the subjects on individual task panels as shown in Figure 1.



FIGURE 1  
FRONT VIEW OF THE PERFORMANCE-TASK PANEL MOUNTED AT  
EACH OF THE FIVE SUBJECT STATIONS

The available data show that all of these tasks possess a sufficiently high reliability to warrant their continued use (2). In addition to study of the effects of multiple-task performances, the effects on watchkeeping performance of additional input information was tested in the present study—specifically studied was a cutaneous signal that was partially redundant with visual information in the Probability Monitoring task. It has been suggested elsewhere (10) that cutaneous stimuli may be used as alerting signals, and they were so used here for Probability Monitoring. With the addition of cutaneous information as an experimental condition, the effect of using more than one kind of sensory input could be studied. Used with the program described earlier the cutaneous signal provided direct empirical evidence on the basis of which assessments could be made (*a*) of its effect on both the time and error aspects of performance of the Probability Monitoring task with which the cutaneous information was redundant and (*b*) of its possible interference (or facilitation) with other tasks.

#### B. METHOD

Two five-man groups of subjects participated in the study; each group worked four hours on each of five consecutive days. The groups were members of AF ROTC units at Emory University and Georgia Institute of Technology. All subjects volunteered to participate and were paid on an hourly basis.

The schedule called for one session (i.e., one day) of orientation and

training followed by four data-taking sessions. There were four program repetitions per session, the task program being a 50 per cent time reduction of that presented in Table 1 (i.e., it consisted of 7.5 minutes of low-demand performance, then of 7.5 minutes of monitoring tasks plus arithmetic, then of 7.5 minutes of monitoring + arithmetic + code-lock solving, etc.). This task program provided an hourly cycling through three levels of performance demand. In addition, it permitted a comparison of results with those of previous studies (2, 3).

The Arithmetic Computation task consisted of the addition of two 3-digit numbers and of the subtraction of a third 3-digit number from the sum of the first two. Problems were presented at a rate of three per minute. Code-Lock Solving was a group-performance task that required discovery of the proper sequential order for depressing five push buttons, one at each subject's work station; intragroup coordination and cooperation was necessary for efficient performance of this task. This was the only task in the battery that was self-paced by the subjects. All other tasks were force paced at specified rates. Target Identification consisted of the presentation of a "target" figure and, after a pause of five seconds, the successive presentation of two comparison figures. The "metric" figures used [see (1)] were composed of square lights from a  $6 \times 6$  matrix. In appearance, they resembled solid bar graphs. Comparison figures had some of the cells randomly changed from their original lit (or unlit) state to the opposite unlit (or lit) state; i.e., the figures were "noisy." The subject's task was to indicate whether the first or the second (or neither) of the comparison figures was the same as the target figure if the noise were to be removed. The three watchkeeping tasks, presented continuously, are described in more detail below.

### 1. *Probability Monitoring*

The four meters across the top of the panel in Figure 1 were used in this task. The pointer on each scale was driven by a random program generator, so that the distribution of momentary pointer settings constituted a normal distribution with a mean at the 12 o'clock position and with a standard deviation of 25 scale units. Nine times during the period of the one-hour program, a bias on one or another of the meters was introduced. The bias caused a shift in the mean of the pointer settings so that the pointer would fluctuate around a mean of 25 scale units to the right or left of the normal 12 o'clock center position. The mean intersignal interval for biases so introduced was 400 seconds; the distribution was that of the right half of a normal distribution with a standard deviation of 200 seconds and a class

interval of 50 seconds. Once a bias was introduced, it remained in the meter until the subject corrected it by moving a switch or until the programmer introduced a new bias into any other meter.

## 2. *Warning-Lights Monitoring*

This task required the subject to detect changes in five green lights normally on and in five red lights normally off. The lights were arranged in pairs of one red and one green on the outer sides of the panels and in the center (see Figure 1). Signal density was approximately nine signals per hour. When either a green light went off or a red light came on, the subject was required to depress the push button under the light in question. When the subject did this, the "nonnormal" condition was corrected; also, his reaction time had been measured to the nearest 0.001 minute.

## 3. *Auditory Vigilance*

The only auditory task used an 1175-cps tone of 0.25-second duration presented with a normal off period of 0.95 second; i.e., the tone was presented once every 1.20 seconds through one earphone of a headset worn by the subject. The subject was required to detect a lengthening of the interval between tones by about 50 per cent; i.e., from 0.95 to 1.40 seconds. Signal density was approximately eight signals per hour.

On two of the four data-taking days, electrical cutaneous stimulation was presented to provide additional information on the Probability Monitoring task. Electrodes of 5-mm diameter were taped to four fingers of the left hand. An indifferent electrode was strapped to the wrist. Electrical cutaneous stimulation (ECS) of 60 cps and 0.5-second duration was presented when a bias was introduced on one of the meters. The purpose was to indicate (a) the moment of occurrence of the bias and (b) the meter on which the bias was introduced—this latter information being indicated by spatial correspondence between the finger stimulated and the four horizontally arranged meters. Thus, stimulation of the little finger indicated a bias on the leftmost meter on the panel; ECS to the ring finger indicated a bias on the second meter from the left, etc.

Although the ECS indicated on which of the four meters a bias should be apparent, it did not indicate the *direction* of the bias (left or right). Thus, the ECS may be said to have provided information that was partially redundant with the visual display in that it did not furnish all of the required information. Group A received ECS during the last two days of data taking. Group B received ECS during the first two days of data taking. Intensity of the

ECS was about 6.0 db relative to the individual subject's absolute threshold; i.e., the intensity was strong enough to yield a distinct tingling sensation, and yet too weak to yield pain. No subject reported emotional or other difficulties when receiving the electrical stimulation.

### C. RESULTS

Performance efficiency for each of the three monitoring and vigilance tasks was assessed as a function of the program section or level of task demand. The performance attained during each section of the program was rank ordered for each measure of each task employed. Statistical significance was tested nonparametrically with  $\chi_r^2$  [see (15, p. 170)]. The results are listed in Table 2. Separate tests were computed for the data obtained with and without the use of the redundant cutaneous information.

When electrical cutaneous stimulation (ECS) was presented, significant variations in performance occurred at the different levels of task demand with Auditory Vigilance and Warning-Lights Monitoring, but *not* with the Probability Monitoring task. The variations in performance of all tasks were significant when the cutaneous signal was not used. It is apparent that performance was most efficient during the task-program subperiods when only the monitoring and vigilance tasks were performed. It was next most efficient when only one additional task was presented, and it was least efficient when the highest level of task demand was used.

The significance of these apparent groupings of performance was assessed by use of a Duncan range test (8). The results were (a) that, at the poorest level, monitoring + targets + code-lock performances were significantly different from monitoring + arithmetic + code-lock performances; (b) that monitoring + arithmetic + code lock was significantly different from the medium-demand levels; (c) that performances with the various tasks at the medium-demand level were not significantly different among themselves; and (d) that performance at the medium level of demand was significantly different from performance with the monitoring tasks only.

During the second of the program sections, Probability Monitoring performance occurring at the end of the program was tested against its performance at the beginning of the program. When ECS was not furnished, performance efficiency did not change significantly from the earlier to the later of these two low-demand sections ( $p > .30$ ). However, when additional information was furnished with the partially redundant cutaneous display, the  $\chi_r^2$  tests indicated that performance during the second low-demand portion of the program was better than that at the first ( $p < .05$ ). Apparently, the



TABLE 2  
PERFORMANCE EFFICIENCY RANKED ACROSS TASK

Task	p	Performance (best to worst)
<i>Sessions with cutaneous stimulation</i>		
Auditory Vigilance	< .001	Monitoring tasks only Monitoring + Target Identification Monitoring + Code Lock Monitoring + Arithmetic Monitoring + Arithmetic + Code Lock Monitoring + Targets + Code Lock
Probability Monitoring (Mean latencies*)	> .10	Monitoring + Code Lock Monitoring + Arithmetic Monitoring tasks only Monitoring + Target Identification Monitoring + Targets + Code Lock Monitoring + Arithmetic + Code Lock
Warning-Lights Monitoring	< .005	Monitoring tasks only** Monitoring + Arithmetic** Monitoring + Code Lock Monitoring + Targets + Code Lock Monitoring + Arithmetic + Code Lock Monitoring + Target Identification
<i>Sessions without cutaneous stimulation</i>		
Auditory Vigilance	< .005	Monitoring tasks only Monitoring + Arithmetic Monitoring + Target Identification Monitoring + Code Lock Monitoring + Targets + Code Lock Monitoring + Arithmetic + Code Lock
Probability Monitoring (Mean latencies*)	< .001	Monitoring tasks only Monitoring + Target Identification Monitoring + Arithmetic Monitoring + Code Lock Monitoring + Arithmetic + Code Lock Monitoring + Targets + Code Lock
Warning-Lights Monitoring	< .001	Monitoring tasks only Monitoring + Arithmetic Monitoring + Code Lock Monitoring + Target Identification Monitoring + Arithmetic + Code Lock Monitoring + Targets + Code Lock

\* Per cent of correct responses indicated results similar to the latency and, therefore, were not analyzed in detail.

\*\* Tied rank.

extra cutaneous stimulation effectively counteracted the "normal" tendencies toward decrements in vigilance performance with time on duty (12).

All three scorable aspects of performance on the Probability Monitoring task were tested to determine the effect of the addition of ECS with each: namely, response latency, per cent of correct responses, and false responses. The mean latencies are shown in Figure 2, and a summary of the results of

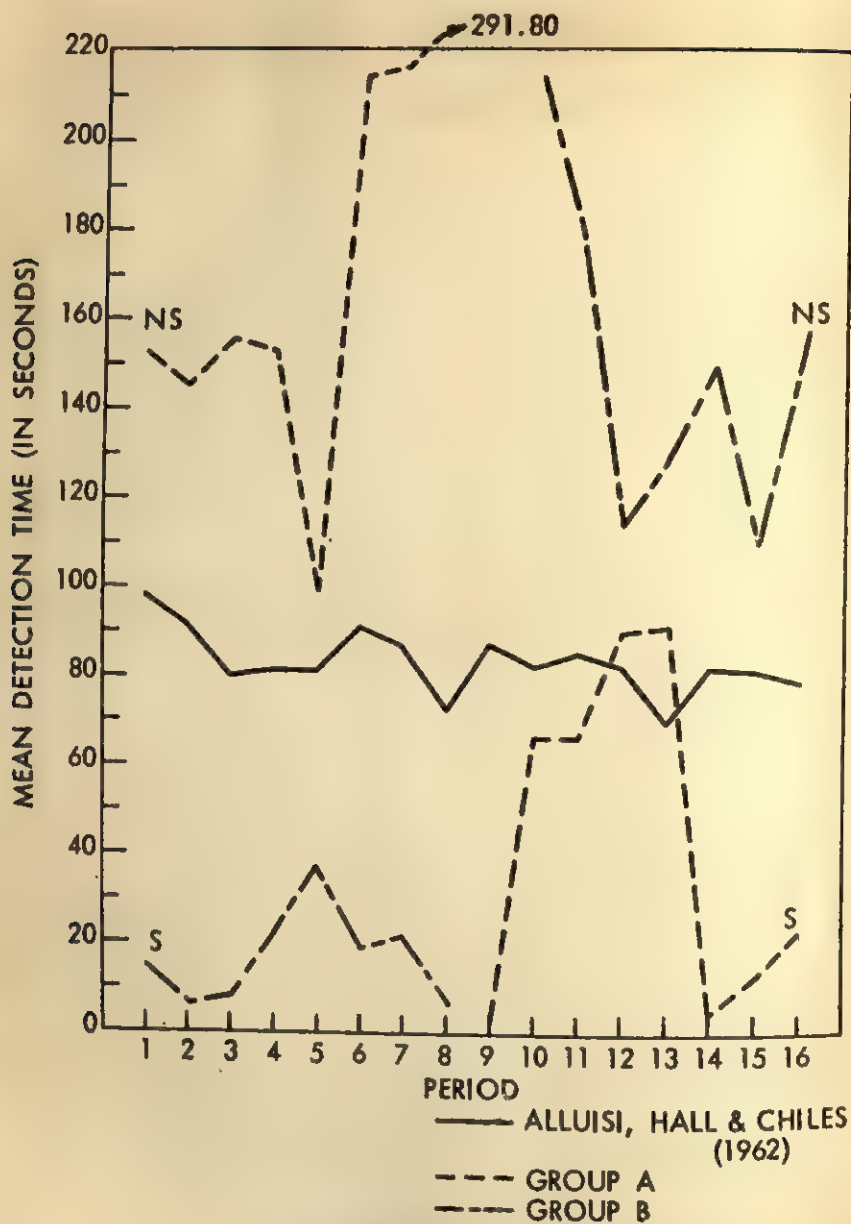


FIGURE 2  
 MEAN RESPONSE LATENCIES TO CORRECT DETECTIONS  
 IN THE PROBABILITY-MONITORING TASK

an analysis of variance of the latencies is provided in Table 3. It is apparent from Table 3 that only one major effect, cutaneous-signal conditions, was significant. That is to say, in terms of response latency, Probability Monitoring performance when ECS was presented was significantly better than when ECS was not used.

TABLE 3  
ANALYSIS OF VARIANCE OF MEAN RESPONSE LATENCIES TO CORRECT DETECTIONS  
IN THE PROBABILITY-MONITORING TASK

Source	df	MS	F	p
Cutaneous-signal conditions (C)	1	703660.50	44.53	< .001
Groups (G)	1	34150.57	—	—
C $\times$ G	1	96.95	—	—
Period (P)	3	4544.51	—	—
C $\times$ P	3	8298.49	—	—
G $\times$ P	3	6643.88	—	—
C $\times$ G $\times$ P	3	2877.49	—	—
Test day (D) (CG)	4	10251.92	—	—
P $\times$ D (CG)	12	9483.28	—	—
Subjects (S) (G)	8	101552.15	—	no test
Residual	120	15801.98	—	no test

The per cent of Probability Monitoring signals correctly detected is shown in Figure 3 for each of the two groups of subjects. The results of an analysis of variance of these per cents yielded a conclusion identical to that implied

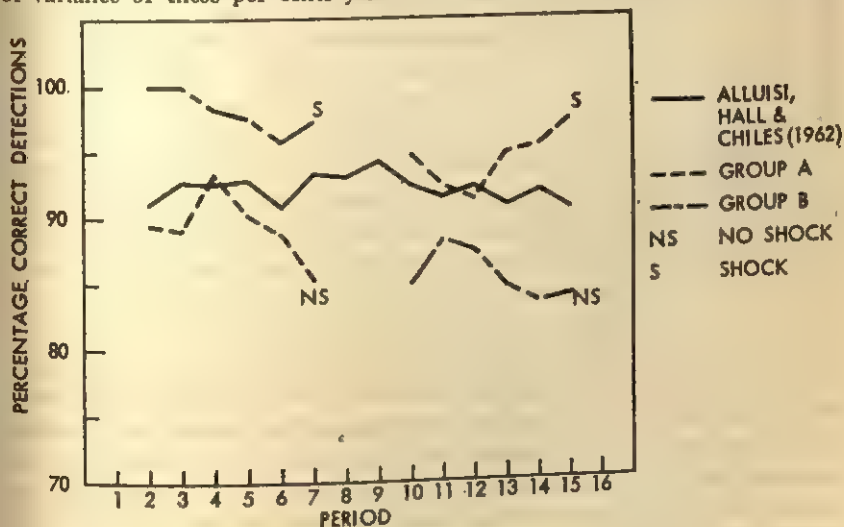


FIGURE 3  
ROLLING MEANS OF PER CENT CORRECT DETECTIONS  
IN THE PROBABILITY-MONITORING TASK

by the latencies: performance with the use of ECS was significantly better ( $p < .001$ ) than when only a visual display was used, and no other main effect or interaction was significant. Significant differences in false responses were not obtained as a function of ECS *versus* no ECS, nor with variations in the level of task demand.

The effects of adding ECS to the Probability Monitoring task on the performances of the other tasks were tested for Code-Lock Solving, Arithmetic Computations, and Target Identifications; the use of ECS has already been shown to have no effect on the performances on the other monitoring and vigilance tasks (*cf.* Table 2). The use of ECS created no significant differences in Code-Lock performances when these occurred only with the monitoring tasks or when they occurred with the monitoring and Target Identification tasks ( $p > .10$  in all cases for each group of subjects). When the Code-Lock Solving occurred with the monitoring and arithmetic tasks, there was no significant difference in performance due to the presentation of ECS (with both groups of subjects considered jointly); however, when considered separately, Group B did better without the use of ECS than with it ( $p < .05$ ). Group B received ECS on the first two days of data taking and not on the last two, whereas Group A was given the reverse order. Thus, this finding may be considered to be a simple practice effect in that it may be expected that efficiency of time-sharing performance during this high-demand portion of the program would improve during the later sessions [*cf.* (3)].

Performance on Arithmetic Computations and Target Identifications did not change as a function of the use of ECS with the Probability Monitoring task, nor did it matter whether these tasks were presented with Code Lock or only with the monitoring tasks ( $p > .90$  in all cases). There were, then, no general effects (detrimental or facilitating) on the performance of other tasks that could be identified as having been caused by the use of electrical cutaneous stimulation with Probability Monitoring.

#### D. DISCUSSION AND CONCLUSIONS

A salient finding of the present study is that monitoring performance varies as a function of the level of task demand. Task demand was systematically varied among low, medium, and high levels in this study; and performance efficiency varied inversely with these levels on all three of the monitoring and vigilance tasks employed (the one exception occurred with the Probability Monitoring task under the condition that paired it with the partially redundant electrical cutaneous stimulation, or ECS). These

changes in the levels of performance efficiency cannot be considered simply a function of time-on-task; performance on all three tasks was most efficient for the low-demand periods, next most efficient during medium demand, and least efficient during high-demand portions of the program (irrespective of the time of occurrence of these levels within the task program).

These variations in vigilance and monitoring performances with different levels of multiple-task activity are of considerable relevance to any attempt to explain watchkeeping behavior (here used as including both vigilance and monitoring activities) with arousal theory [for a review of such attempts, see (9)]. In brief, these explanations state essentially that the greater the total volume of stimulus input per unit of time, the greater will be the efficiency of detection of critical signals in a watchkeeping task. The results of our study are precisely in opposition to this prediction; *viz.*, performance was most efficient during low-demand, next most efficient for medium-demand, and least efficient with high-demand levels, where these levels were associated with comparably different levels of stimulus input and comparably different levels of multiple-task activity.

It is obvious that arousal theory is not a good predictor of performance in the kind of situation used in this study. If it is useful at all, arousal theory may be able to predict the levels of performance efficiency where the total volume of stimulus input is extremely low; *i.e.*, under extremely "deprived" conditions only. Once some minimal number of stimuli per unit of time has been achieved, then for anything above such a low level, arousal theory should probably be disregarded in favor of some other explanation of watchkeeping behavior.

It is interesting to note that the addition of ECS to the Probability Monitoring task resulted in an improvement in performance efficiency during the second low-demand period (at the end of the task program) as compared to the low-demand period at the beginning of the program ( $p < .05$ ). It is doubtful that this improvement may be ascribed to any simple learning processes, for there was no significant change in efficiency between these two periods without ECS ( $p > .10$ ); also, the general level of efficiency was significantly better with the cutaneous information than without it. It is suggested that the results are consistent with the findings of Buckner, Hara-bedian, and McGrath (7), who reported that "good" observers in a vigilance task (defined as the upper 27 per cent in performance efficiency) maintain their efficiency well and often improve with time-on-task, whereas "poor" observers (the lower 27 per cent) demonstrate rather severe performance decrements under the same experimental conditions. That is to say, the addi-



tion of the cutaneous stimulation as partially redundant information had the effect of making the observers "better" performers, not necessarily through the learning route, but perhaps through use of a more nearly optimal method of presenting the critical information.

The use of electrical cutaneous stimuli as partially redundant Probability Monitoring signals yielded several other important results: (a) there were significant improvements ( $p < .001$ ) in Probability Monitoring performances in terms of both the per cent of correct responses and response latencies; (b) Probability Monitoring performance did *not* change significantly ( $p > .10$ ) as a function of the level of task demand as it did without ECS and as all other monitoring tasks changed; (c) cutaneous stimuli had no significant effect on the performance of other tasks, these including such tasks as additional monitoring and vigilance displays, mental arithmetic, group coordination and cooperation, and the perception of visual form.

In other words, it seems reasonable to conclude that the use of electrical cutaneous stimulation as a partially redundant warning or alerting signal served to improve the performance of the specific monitoring task to which it was added and to maintain the performance of that task at a high level, in spite of variations in the level of task demand. Furthermore, it achieved these results without any loss of performance efficiency in the other tasks.

The most important implication of these results is that cutaneous communications may not necessarily interfere with the processing of other information received with auditory and visual displays. This conclusion warrants continued and increased efforts to devise efficient cutaneous communication systems that can be used in transmitting various kinds of information at different levels of complexity. With proper encoding, such systems could be used to add redundant or partially redundant information to other displays or to provide a primary information display. The extent to which the real limitation—the "total information-processing capacity" of man—will permit performance gains from use of cutaneous communication in highly complex situations is yet to be determined; as Miller (14) has pointed out, man appears to be able to handle more and more information when he is able to encode it efficiently in "chunks" rather than in single units.

#### E. SUMMARY

An investigation of performance efficiency in several watchkeeping tasks was made with use of a task program encompassing various levels of task demand and several different kinds of activity. The additional tasks involved

mental arithmetic, group coordination and cooperation, and form perception systematically presented during selected portions of the task program.

The results indicated statistically different watchkeeping performances as a function of the level of task demand. Efficiency was highest when the task demand was low (and stimulus input was also relatively low), next most efficient with medium-level demand, and least efficient with high-level task demand. The addition of electrical cutaneous stimulation as partially redundant information in a monitoring task: (a) improved performance in that task, (b) served to maintain efficiency in spite of variations in the level of task demand, and (c) did not interfere with performance efficiency on other (concurrent) tasks.

It is concluded that arousal theory is not a good predictor of watchkeeping performance in the kind of situation studied and may apply only in "stimulus-deprived" conditions.

#### REFERENCES

1. ALLUISE, E. A. On the use of information measures in studies of form perception. *Percept. & Mot. Skills*, 1960, **11**, 195-203.
2. ALLUISE, E. A., HALL, T. J., & CHILES, W. D. Group performance during four-hour periods of confinement. AMD Rep. No. MRL-TDR-62-70, Wright-Patterson AFB, Dayton, Ohio, 1962.
3. ALLUISE, E. A., CHILES, W. D., HALL, T. J., & HAWKES, G. R. Human group performance during confinement. AMD Rep. No. MRL-TDR-63-87, Wright-Patterson AFB, Dayton, Ohio, 1963.
4. BAKER, R. A., WARE, J. R., & SIPOWICZ, R. R. Vigilance: A comparison in auditory, visual, and combined audio-visual tasks. *Can. J. Psychol.*, 1962, **16**, 192-198.
5. BINFORD, M. R., & LOEB, M. Arousal and interference effects of a concurrently performed auditory task on the detection of threshold visual signals. Paper read to Southern Society for Philosophy and Psychology, Miami Beach, Fla., April, 1963.
6. BUCKNER, D. N., & McGRATH, J. J. A comparison of performances on single and dual sensory mode vigilance tasks. Tech. Rep. 8, Human Factors Res., Inc., Los Angeles, Calif., February, 1961.
7. BUCKNER, D. N., HARABEDIAN, A., & McGRATH, J. J. A study of individual differences in vigilance performance. Tech. Rep. 2, Human Factors Res., Inc., Los Angeles, Calif., January, 1960.
8. DUNCAN, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1-7.
9. FRANKMANN, J. P., & ADAMS, J. A. Theories of vigilance. *Psychol. Bull.*, 1962, **59**, 257-272.
10. HAWKES, G. R. Potential answers to communication problems. *Aerospace Med.*, 1962, **33**, 657-662.
11. LOEB, M., & HAWKES, G. R. Detection of differences in duration of acoustic and electrical cutaneous stimuli in a vigilance task. *J. of Psychol.*, 1962, **54**, 101-111.
12. MACKWORTH, N. H. Researches on the Measurement of Human Performance. London: H. M. Stationery Office, 1950.

13. McGRATH, J. J. Performance sharing in dual-mode monitoring. Tech. Rep. 740-1, Human Factors Res., Inc., Los Angeles, Calif., January, 1962.
14. MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81-97.
15. SIEGEL, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.

*U. S. Army Medical Research and Development Command  
Office of The Surgeon General  
Washington, D. C. 20315*

## MULTITHERAPIST INTERVIEWS UTILIZING LSD \*

*South Oaks Research Foundation, Amityville, Long Island, New York*

---

A. ROLO, L. W. KRINSKY, H. A. ABRAMSON, AND L. GOLDFARB

---

In the last few years, there has been increasing use of LSD as an adjunct to psychotherapy. Some investigators (5, 6) utilize the drug primarily for its transcendental effect, sometimes leaving the patient in a darkened room with music. Others (3) rely on the cathartic effect of the drug and allow the patient to remain alone or with a nurse. The patient then speaks into a tape recorder. Still others (1, 2) rely on a standardized interview technique on a one-to-one basis or on a group-therapy technique (7). In summary, the nature of the therapeutic process is as varied as that of the therapists.

Utilizing the above procedures, we have found that patients do not develop the degree of insight plus motivation to continue in intensive therapy. In follow-up studies of patients who had had LSD once or twice without after care, results have been very poor (very high relapse rate).

The following technique was developed as a modification of previously described work (4). The major purpose was to devise a technique in which previously untapped anxiety-laden areas were probed (most of the patients had been in therapy previously). In addition, it was hoped that these patients could be motivated to continue in a therapeutic relationship following the LSD session.

At the South Oaks Psychiatric Hospital, it has been found that multitherapist interviews are practical and maintain a patient as an active participant in the therapeutic process. The multitherapist approach has been employed in the past, but apparently not under conditions in which the patient was administered LSD as an adjunct to psychotherapy.

Our multitherapist extended interviews begin anywhere from 8:00 to 8:30 A.M. The patient is administered 100 micrograms of LSD (supplied by Sandoz, Inc.) orally. This is administered with one of the interviewers present. Within 30 to 45 minutes, the remaining two interviewers join the conference.

As the effect of the drug heightens, the patient reacts to the presence of therapists. Differing transference reactions to each therapist are made

---

\* Received in the Editorial Office on April 20, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

more manifest. Almost invariably one therapist is selected as a "parent figure" or "protector"; another, as the "enemy," a therapist whom the patient evidently feels safe in attacking. The third therapist is variously identified as contemporary: a teacher, or a professor.

The initial phases of the conference are conducted vis-à-vis, with the patient sitting in an undarkened room. Within one to one and one-half hours, as the reaction heightens, the patient is encouraged to lie in bed and to associate to the material previously discussed. At that time, the room is darkened.

Some of the patients assume a foetal position. Others seem to enter deeply into an almost hypnotic state by repeating one statement over and over. During this phase of the conference, the orientation of the therapists moves from a more directive to a less directive approach. The patient frequently moves spontaneously into highly traumatic childhood experiences and may hyperventilate or act out certain feelings. None of this acting out interferes with therapist interviews. More success has been encountered with patients who have previously been in analysis with unsuccessful results as contrasted with groups of relatively psychologically unsophisticated patients.

After the patient has entered into a heightened reaction, such a reaction is permitted to continue relatively unabated for one to one and one-half hours. On occasion, the reaction is terminated when the patient begins to hyperventilate excessively or begins to show signs of dealing with material that would overcome ego reserves.

Another aspect of the procedure that is considered very important consists of having lunch with the patient as a device for terminating the interview. Associating on such an intimate level with three persons whom the patient also sees more or less as authority figures invariably has some beneficial effect. The patient occasionally is able to begin to resolve some distortions as to authority figures. In addition, this period is used to recapitulate some of the findings. The patient continues as an active participant. He is often so involved as to be unable to eat.

Previously described studies utilize the administration of 25 to 400 micrograms of LSD. Our experience is that very high dosages serve to inhibit rather than enhance the response pattern during psychoanalytically oriented interviews. The usual dosage is 100 micrograms administered orally after breakfast and approximately one hour before the conference is to begin. If there is an insufficient response, an additional 25 to 75 micrograms is administered hypodermically. The dose rarely exceeds 200 micrograms.

It is our experience that conferences can profitably be terminated from four



and one-half hours to five hours after beginning. The patient is then administered one of the phenothiazines or a barbiturate.

In multitherapist interviews utilizing 47 patients over a three-year period, 27 patients continued in intensive therapy (58 per cent). Forty (85 per cent) of these patients expressed a belief that the interview procedure had been of benefit. Staff opinion was that 30 patients (64 per cent) had benefited. Research is continuing in follow-up studies of these patients.

#### REFERENCES

1. ABRAMSON, H. A. Lysergic acid diethylamide (LSD-25): III. As an adjunct to psychotherapy with elimination of fear of homosexuality. *J. of Psychol.*, 1955, **39**, 127-155.
2. ABRAMSON, H. A., Ed. *The Use of LSD in Psychotherapy*. New York: Josiah Macy, Jr., Foundation, 1960.
3. FREDERKING, W. Intoxicant drugs (mescaline and lysergic acid diethylamide) in psychotherapy. *J. Nerv. & Ment. Dis.*, 1955, **121**, 262-266.
4. ROLO, A., KRINSKY, L. W., & GOLDFARB, L. LSD as an adjunct to psychotherapy with alcoholics. *J. of Psychol.*, 1960, **50**, 85-104.
5. SAVAGE, C., TERRILL, J., & JACKSON, D. D. LSD, transcendence, and the new beginning. *J. Nerv. & Ment. Dis.*, 1962, **135**, 425-439.
6. SHERWOOD, J. N., STOLAROFF, M. J., & HARMAN, W. W. The psychedelic experience—A new concept in psychotherapy. *J. Neuropsychiat.*, 1962, **1**, 370-375.
7. TENENBAUM, B. Group therapy with LSD-25. *Dis. Nerv. Syst.*, 1961, **22**, 459-462.

*South Oaks Psychiatric Hospital*

*The Long Island Home, Ltd.*

*Sunrise Highway*

*Amityville, Long Island, New York*

\$20.00 per annum  
\$15.00 per volume  
\$7.50 per half volume

QUARTERLY  
Two volumes per year  
Immediate publication

October, 1964  
Volume 58, Second Half

(Founded in 1935 by Carl Murchison)

# The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

OCTOBER, 1964

(Manuscripts are printed in the order of final acceptance)

- Stimulus and motivational determinants in temporal perception . . . . . 243  
By JOEL S. WARM, LEWIS F. GREENBERG, AND C. STUART DUBE, II
- Short-term retention as a function of item frequency . . . . . 249  
By KENNETH E. LLOYD
- Participation, interaction and role behavior in groups selected from the extremes of the open-closed cognitive continuum . . . . . 255  
By SALVATORE V. ZAGONA AND LOUIS A. ZURCHER, JR.
- Chemical studies of behavior: VIII. Some anomalous onset effects of benzquinamide, a "tranquilizer" . . . . . 265  
By JEAN S. CAMERON AND G. R. WENDT
- ✓ A further analysis of the role of creative thinking in high-school achievement . . . . . 277  
By KAORU YAMAMOTO

(OVER)

Copyright, 1964, by The Journal Press  
Provincetown, Massachusetts, U. S. A.

All rights reserved

Published at 35B New St., Worcester, Mass.  
Second-class postage paid at Worcester, Mass.

✓ Evaluation of some creativity measures in a high school with peer nominations as criteria . . . . .	285
BY KAORU YAMAMOTO	
Code learning and color . . . . .	295
BY C. W. CRANNELL	
Mean associative latencies of 200 CVC trigrams . . . . .	301
BY RONALD C. JOHNSON	
The sociometric composition of the group as a determinant of yielding to a distorted norm . . . . .	307
BY FRANK B. W. HARPER AND READ D. TUDDENHAM	
Consistency of cross-test individual differences in personality-inventory faking BY JOHN R. BRAUN	313
Shifts in connotative meaning of words as a function of context and rein- forcement . . . . .	317
BY M. E. TRESSELT AND M. S. MAYZNER	
Critical flicker frequency for high pulse-to-cycle fractions and high intensity BY JANE E. RANNEY AND S. HOWARD BARTLEY	329
CFF for short trains of photic stimulation having various temporal distribu- tions and separations . . . . .	333
BY THOMAS M. NELSON, S. HOWARD BARTLEY, AND EARL S. HARPER	
The effects of cycling upon CFF when stimulus purity is varied . . . . .	343
BY THOMAS M. NELSON, S. HOWARD BARTLEY, AND FRED C. BLECK	
Subjective color: A new method of producing the phenomenon . . . . .	353
BY CHARLES A. KNEHR AND ROBERT J. LORENZ	
Trait judgments of military pilots from photographs . . . . .	357
BY D. J. BAER	
Conflict models, games, and drinking patterns . . . . .	361
BY HENRY S. G. CUTTER	
Item context and position effects on multiple-choice tests . . . . .	369
BY LEWIS R. AIKEN, JR.	
Aesthetic preferences of art students and mathematics students . . . . .	375
BY RUSSELL EISENMAN AND SANDRA COFFEE	
Description of a central mechanism related to diverse brightness phenomena . . . . .	379
BY THOMAS M. NELSON AND S. HOWARD BARTLEY	
Drugs and judgment: Effects of amphetamine and secobarbital on self- evaluation . . . . .	397
BY GENE MARSHALL SMITH AND HENRY K. BEECHER	
Behavior in the attainment of economic concepts: II. . . . .	407
BY RACHEL S. SUTTON	
Stepwise discriminant-function analysis of seminary-candidate MMPI scores . . . . .	413
BY DANIEL J. BAER AND JAMES F. MOYNIHAN, S.J.	
Human-figure drawings in relation to modernization in Sudan . . . . .	421
BY MALIK B. BADRI AND WAYNE DENNIS	
The mood effects of music: A comparison of data from four investigators . . . . .	427
BY MELVIN G. RIGG	
An individual action-research method for intensive exploration of adaptive difficulties in the prepuberty stage . . . . .	439
BY C. J. MORGAN	
Ego identity and some aspects of personal adjustment . . . . .	459
BY S. M. HOWARD AND J. F. KUBIS	
Set and shock-stress effects upon illusion perception . . . . .	467
BY DANIEL J. BAER	

## STIMULUS AND MOTIVATIONAL DETERMINANTS IN TEMPORAL PERCEPTION\*

*Department of Psychology, University of Bridgeport*

JOEL S. WARM, LEWIS F. GREENBERG, AND C. STUART DUBE, II

### A. INTRODUCTION

Marginal stimulus situations are typically employed in assessing the influence of personal determinants (e.g., needs and interests) on perception (1). According to Falk and Bindra (8), temporal judgments may provide a useful dimension in this regard because such judgments are based on meager cues. Previous research concerning the influence of motivation on temporal experience has focused primarily upon experimentally induced motives (8, 9, 12, 14, 17, 18, 23), and few data are available regarding the effects of more implicit aspects of individual dynamics upon the perception of time (10, 27). Accordingly, the present study will attempt to determine if characteristic changes can be identified in the perceived duration of briefly presented visual stimuli as a function of one such aspect of individual organization: viz., personal values.

In an early study of the sensitizing effect of motivational factors on perception, Postman, Bruner, and McGinnies (21)—using the Allport-Vernon Study of Values (2)—found that recognition thresholds for tachistoscopically presented words varied inversely with the value areas of the *O*. Later findings have demonstrated, however, that lower recognition thresholds could be attributed to a response probability factor based upon *O*'s familiarity with the words employed (*cf.* 20). Inasmuch as temporal experience seems highly sensitive to variations in stimulus conditions (5, 10), word familiarity may have important implications for the study of perceived *duration* of brief visual forms as well as for recognition thresholds of these stimuli. Hence, in the present design, perceived duration will be examined in relation to stimulus familiarity as well as in relation to *O*'s personal values. Such an approach might offer some specification of a crucial stimulus factor in temporal experience and may have relevance to the stimulus-*vs.*-subjective determinant controversy in perceptual theory (22, 29).

\* Received in the Editorial Office on April 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

### B. PROCEDURE

Forty-five *O*'s, 26 men and 19 women, were drawn from students in introductory psychology at the University of Bridgeport. The experimental laboratory consisted of two adjacent rooms set up as projection and viewing stations. A Graflex School Master 500 slide projector, with a Lafayette All-Purpose Tachistoscope Attachment, was used to present a series of 24 slides, each slide containing a stimulus word typed in capital letters. The stimuli were flashed from the projection station through a one-way vision window onto a screen in the viewing station. All the stimuli were presented for *one* second with an intertrial interval of 15 seconds. An intercom system allowed for one- or two-way communication at *E*'s discretion, and light coming from the projection room permitted *O* to remain light-adapted throughout the experimental session.

Two words representing each value category of the Allport-Vernon Study of Values (2) were selected from a list presented by Solomon and Howes (25). The Lorge-Thorndike word counts (26) were used to establish the relative frequency of occurrence of the stimulus words in ordinary usage and hence their familiarity to *O* (*cf.* 25). In determining word frequency, the average of the data for each word from the Lorge magazine count and the Lorge-Thorndike semantic count was obtained after a procedure recommended by Howes and Solomon (13) and by McGinnies, Comer, and Lacey (16). The mean frequency per 4-1/2 million words for stimuli dichotomized as frequent and infrequent was 555.23 and 14.08 respectively. Stimulus words classified as to frequency and value are listed in Table 1.

*Os* were tested in two groups of 22 and 23 each. They were told that a series of words would be flashed on the screen in front of them and that they were to record each word and their impression of its flash duration on answer forms provided. *Os* thus responded by the method of verbal estimation (4). Timepieces were made inaccessible before the experiment began, and *O* was instructed not to count or tap while the stimulus was on. One *E* operated the projection equipment while the other remained in the viewing station with the *Os* to prevent intersubject communication. *Os* were not given information regarding the accuracy of their judgments, and the order of stimulus presentation was randomized for each group. The Allport, Vernon, Lindzey Study of Values (3) was administered at the close of each session.

### C. RESULTS

Average temporal judgments for all stimulus words were determined for each *O*. Mean judgments for the two groups tested under different stimulus



TABLE 1  
STIMULUS WORDS

Value	Frequent	Infrequent
Theoretical	scientific physics	percipience statics
Economic	automobile market	frugality assets
Political	government lawyer	assiduous barrister
Esthetic	orchestra poetry	vignette etcher
Religious	religious churches	celestial theistic
Social	hospitality service	condolence benign

presentation orders were .96 and .80 second. This difference was not statistically reliable ( $t = 1.14$ ,  $df = 43$ ); therefore data from both groups of *O*s were pooled in all comparisons.

Value ranks for each *O* were determined from his scores on the Allport, Vernon, Lindzey scale. The highest value area was ranked as one and the lowest as six. Average temporal judgments for words representing each value rank were determined for each *O*, and mean judgments across *O*s for frequent and infrequent words are presented in Table 2.

TABLE 2  
MEAN TEMPORAL JUDGMENTS (IN SECONDS) AS A FUNCTION OF VALUE RANK

Stimulus category	Value ranks					
	1	2	3	4	5	6
Frequent words	.93	.94	1.01	.92	.89	.91
Infrequent words	.84	.81	.85	.78	.80	.87

Generally, there appears to be no consistent tendency for temporal judgments to vary systematically with the value rank. An analysis of variance of these data (6)—shown in Table 3—indicates that, while *O*s differed significantly among themselves, there was no significant main effect for personal values.

To examine further the relationship between word value and temporal judgment, a separate analysis of variance was performed on the data for 20 *O*s with distinctively high departures from the standardized norms for any given area on the value scale (3). Results paralleled those for the total sample; i.e., although there were significant differences between *O*s, there was no consistent relationship between value rank and perceived duration.

TABLE 3  
ANALYSIS OF VARIANCE OF TEMPORAL JUDGMENTS AS A FUNCTION OF VALUE RANK

Source of variation	df	Frequent words		Infrequent words	
		Mean square	F	Mean square	F
Between subjects	44	1.4817	27.96*	1.2874	25.34*
Between values	5	.0777	1.47	1.0630	1.06
Error	220	.0530		.0508	
Total	269				

\*  $p < .01$ .

Further examination of Table 2 reveals a uniform tendency for temporal judgments to vary directly as a function of stimulus familiarity. Mean judgments across *O*s for all frequent and infrequent words were .93 and .83 second respectively; a *t* test for correlated samples indicates that this difference is significant at the .01 level ( $t = 4.54$ ,  $df = 44$ ).

#### D. DISCUSSION

Results of the present experiment demonstrate that stimulus familiarity is of crucial significance not only in determining recognition thresholds for tachistoscopic words (*cf.* 20), but also in determining their perceived *duration*. Additionally, significant differences between subjects are consistent with earlier studies illustrating wide individual variation in the perception of time (*cf.* 7, 11, 15, 28).

The tendency for temporal experience to vary directly with word frequency may perhaps be accounted for by Fraisse's notion (10) that perceived duration varies directly with the amount of change experienced by *O* in his ambient environment. According to this position, direction of attention to the passage of time itself is a primary means of increasing the rate of perceived change. Considering the stimulus conditions of the present study in light of previous research, it seems plausible to assume that frequent words could be identified more rapidly than infrequent ones (20). Consequently, more of *O*'s attention during the one-second presentation interval could be directed to time for the former than for the latter, resulting in a longer experience of time for the frequent stimuli.

Lack of a significant relationship between temporal judgment and stimulus value accords with the nativist position (*cf.* 22, 29) that perception can be accounted for in terms of stimulus-response correlates and that nonstimulus explanations are unnecessary. However, a direct relation between temporal judgment and word frequency is not consistent with such attempts to negate experiential influences in visual perception.

Failure of temporal judgments to vary systematically with personal value is generally inconsistent with the work of Pearl and Berg (19) and of Siegman (24), who found significant relations between personality factors and time estimation. These studies utilized thematic pictures and open-interval stimuli respectively, presented for periods of five to 30 seconds. Such conditions are less structured than those of the present study. Further, these writers worked with conflict and anxiety factors respectively, while personal values were used herein. It appears that there may be need for theoretical and empirical specification of both the nature of the stimulus situation presented to *O* and the dimension of personality measured when studying the influence of personal variables on the perception of time.

### E. SUMMARY

An attempt was made to determine the relative influence of stimulus and motivational determinants in the perceived duration of tachistoscopic words. Forty-five *O*'s judged the duration of 24 words related to the categories of the Allport, Vernon, Lindzey Study of Values. Stimulus words were dichotomized with respect to their relative frequency of usage and their familiarity to *O*. Results indicated a direct relationship between perceived duration and stimulus familiarity, but no significant relationship between temporal judgment and personal value.

### REFERENCES

1. ALLPORT, F. H. *Theories of Perception and the Concept of Structure*. New York: Wiley, 1955.
2. ALLPORT, G. W., & VERNON, P. E. *A Study of Values*. Boston: Houghton-Mifflin, 1931.
3. ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of Values* (3rd ed.) Boston: Houghton-Mifflin, 1960.
4. BINDRA, D., & WAKSBERG, H. Methods and terminology in studies in time estimation. *Psychol. Bull.*, 1956, **53**, 155-159.
5. DOEHRING, D. G. Accuracy and consistency of time-estimation by four methods of reproduction. *Amer. J. Psychol.*, 1961, **74**, 27-35.
6. EDWARDS, A. L. *Experimental Design in Psychological Research*. New York: Rinehart, 1950.
7. ESON, M. E., & KAFKA, J. S. Diagnostic implications of a study in time perception. *J. Gen. Psychol.*, 1952, **46**, 169-183.
8. FALK, J. L., & BINDRA, D. Judgment of time as a function of serial position and stress. *J. Exper. Psychol.*, 1954, **47**, 279-282.
9. FILER, R. J., & MEALS, D. W. The effects of motivating conditions on the estimation of time. *J. Exper. Psychol.*, 1949, **39**, 327-331.
10. FRAISSE, P. *The Psychology of Time*. New York: Harper & Row, 1963.
11. HARTON, J. J. The relation of time estimates to actual time. *J. Gen. Psychol.*, 1939, **21**, 219-224.

12. HINDLE, H. M. Time estimates as a function of distance traveled and relative clarity of a goal. *J. Personal.*, 1951, **19**, 483-490.
13. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. Exper. Psychol.*, 1951, **41**, 401-410.
14. LANGER, J., WAPNER, S., & WERNER, H. The effect of danger upon the experience of time. *Amer. J. Psychol.*, 1961, **74**, 94-97.
15. LOEHLIN, J. C. The influence of different activities on the apparent length of time. *Psychol. Monog.*, 1959, **73**, 1-27.
16. MCGINNIES, E., COMER, P. B., & LACEY, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. Exper. Psychol.*, 1952, **44**, 65-69.
17. MEADE, R. D. Time perception as affected by need tension. *J. of Psychol.*, 1960, **49**, 249-253.
18. ———. Effect of motivation and progress on the estimation of longer time intervals. *J. Exper. Psychol.*, 1963, **65**, 564-567.
19. PEARL, D., & BERG, P. S. D. Time perception and conflict arousal in schizophrenia. *J. Abn. & Soc. Psychol.*, 1963, **66**, 332-338.
20. PIERCE, J. Determinants of threshold for form. *Psychol. Bull.*, 1963, **60**, 391-407.
21. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. Abn. & Soc. Psychol.*, 1948, **43**, 142-154.
22. PRATT, C. C. The role of past experience in visual perception. *J. of Psychol.*, 1950, **30**, 85-107.
23. ROSENZWEIG, S., & KOHT, A. C. The experience of duration as affected by need tension. *J. Exper. Psychol.*, 1933, **16**, 745-774.
24. SIEGMAN, A. W. Anxiety, impulse control, intelligence, and the estimation of time. *J. Clin. Psychol.*, 1962, **28**, 103-105.
25. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.
26. THORNDIKE, E. L., & LORGE, I. The Teachers Word Book of 30,000 Words. New York: Teachers College, Columbia University, 1944.
27. WALLACE, M., & RABIN, A. I. Temporal experience. *Psychol. Bull.*, 1960, **57**, 213-236.
28. YOUNG, S. A., & SUMNER, F. C. Personal equation, frame of reference, and other observations in remote reproductions of filled and unfilled time intervals. *J. Gen. Psychol.*, 1954, **51**, 333-337.
29. ZUCKERMAN, C. B., & ROCK, I. A reappraisal of the roles of past experience and innate organizing processes in visual perception. *Psychol. Bull.*, 1957, **54**, 269-296.

Department of Psychology  
University of Louisville  
Louisville, Kentucky

## SHORT-TERM RETENTION AS A FUNCTION OF ITEM FREQUENCY\*<sup>1</sup>

*Department of Psychology, Washington State University*

---

KENNETH E. LLOYD

---

### A. INTRODUCTION

This study investigated the effect on recall of varying the number of times items to be stored were presented to subjects (*Ss*) each day. This method of varying item frequency may be contrasted with the design that employs as items words that differ in their English text frequency, but the experimental frequency of which is held constant (e.g., 1, 5). In the present method, *Ss* had to learn which items occurred with high or low frequencies. Once learned it was expected that, when uncertain about a recall, *Ss* would guess high-frequency items and thereby increase the likelihood of a correct recall. However, this strategy could result in fewer correct recalls of low-frequency items. The experiment was designed to provide a check on this possibility.

Two groups were compared. They differed in the range of item frequencies presented to them. Items in the high-range group (Group HR) varied from one to 16 occurrences per day; items in the narrow-range group (Group NR) occurred with frequencies between one and five. Items were divided into three frequency classes. Class I items had a high frequency in Group HR and a medium frequency in Group NR. Class I items were expected to produce fewer errors in Group HR than in Group NR. Class II items had the same, but relatively low, frequency in both groups. These items provided the check on whether *Ss* in Group HR recalled high-frequency items at the expense of low-frequency items. Class III items had a low frequency in Group HR and a medium frequency in Group NR. Class III items were required to fill out a total sequence of items; no predictions were made about their recall scores.

The sequential memory task (2, 3) was used. This task requires *Ss* to store and recall a long series of tape-recorded items. Sequences are constructed so that the average number of items being stored (the average storage load) changes from time to time. In the present study, two average storage loads

---

\* Received in the Editorial Office on May 4, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by the United States Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49(638)-805. Much of the work reported here was performed by William A. Johnston.



(ASL) were used in order to learn if the effects of item frequency varied as the storage requirements of the task changed.

### B. METHOD

Each sequence was composed of eight parts (subsequences) constructed so that their ASL was either 3.5 or 6.5 items. Four sequences were recorded to counterbalance partially the serial order of the subsequences and of the words. The individual storage loads making up the ASLs were 2, 3, 4, and 5 for the 3.5 ASL and 5, 6, 7, and 8 for the 6.5 ASL. These individual loads occurred in a wide variety of serial orders. Each subsequence contained either 13 or 14 recall points. In order to maintain the desired loads near the ends of the subsequences, one to five extra items were inserted. These were called for immediately after the 13th or 14th recall point and were not scored. For each recall point two items had been previously presented; thus *S* always recalled two items at each recall point. A sequence contained a total of 240 items and 120 recall points.

Sixty English words, judged to be unrelated, were arbitrarily grouped into six-letter categories of 10 words each. On Day 1, *Ss* were given a list of these words grouped into their letter categories. When presented for storage, each word was always paired with its appropriate letter (e.g., L-roof). Later in the sequence the letter, L, occurred alone as a cue for *S* to recall "roof." Items were recorded at a 4-second rate and recall points at a 3-second rate. A 500-cps tone preceded each recall point to aid *S* in detecting them.

The six letters occurred equally often in all sequences. Only the words varied in frequency. The frequency range was 16 to 1 for Group HR and 5 to 1 for Group NR. Item frequency was based on occurrences per ASL per day; a word with a frequency of 16 occurred 16 times a day at the 3.5 load and 16 times a day at the 6.5 load. Each frequency was replicated six times: i.e., once in every letter category. The frequencies for Class I items were 16 and 8 for Group HR and 4 and 4 for Group NR. The frequencies for Class II items were 4, 2 and 1 for both groups. The frequencies for Class III items were 1 and 2 for Group HR and 5 for Group NR. Class III included those items that were used both in the subsequence and as extra items at the ends of subsequences. The placement of a word in a sequence was determined not only by the ASL of a particular subsequence, but also by the individual storage load for a given recall point. Thus a word that occurred 16 times per sequence was placed twice at each of the four individual loads comprising the 3.5 ASL and twice at those comprising the 6.5 ASL. Words were randomly assigned their item frequency.

The 20 *Ss*, Washington State University undergraduates, were alternately assigned to the two groups. The *Ss* were reimbursed for their services, and were instructed to remember each letter-word pair as they heard it on the tape and to recall the words when the appropriate letter was heard. No instructions were given regarding either independent variable. Each *S* responded to two sequences a day for four days. The *E* recorded all responses. An error was defined as a failure to respond or as a recall of a word not requested at that recall point.

### C. RESULTS

#### 1. Total Errors

All *Ss* on each day responded to eight subsequences at each value of ASL. The means of errors over these repeated measures were used in the analyses and are in Table 1. An analysis of variance of recall errors on the first sequence on Day 1 indicated Range effect was not significant ( $F = 2.36$ ;  $df = 1, 18$ ;  $p > .05$ ). ASL was significant ( $F = 98.89$ ;  $df = 1, 18$ ;  $p < .01$ ), but the Range  $\times$  Load interaction was not ( $F = 2.52$ ;  $df = 1, 18$ ;  $p > .05$ ). The groups did not differ in total errors at the beginning of the study.

TABLE 1  
MEAN RECALL ERRORS FOR THE TWO GROUPS AT BOTH ASLS AND AT DIFFERENT STAGES OF PRACTICE

ASL	Group	Day 1		Days 1-4		Day 4	
		Mean	SD	Mean	SD	Mean	SD
3.5	HR	12.50	3.32	8.27	2.95	5.67	3.38
	NR	13.40	2.87	11.24	3.38	9.65	4.57
6.5	HR	17.50	2.94	14.24	3.05	12.05	3.83
	NR	20.30	2.15	18.70	2.04	17.20	2.71

The Range effect was significant over Days 1 through 4 ( $F = 8.02$ ;  $df = 1, 18$ ;  $p < .02$ ) and on Day 4 ( $F = 42.50$ ;  $df = 1, 18$ ;  $p < .01$ ). The ASL effect was significant ( $p < .01$ ) on Days 1-4 ( $F = 305.34$ ;  $df = 1, 18$ ) and on Day 4 ( $F = 170.19$ ;  $df = 1, 18$ ). The Range  $\times$  ASL interaction was not significant. Group HR consistently recalled more items than Group NR. An analysis of errors at the frequency classes was conducted next to pinpoint the locus of this facilitation of recall for Group HR.

#### 2. Frequency Errors

The errors for all words in a given frequency class were tallied and a per cent error computed. Mean per cent errors for Day 1 (Sequence 1) and Day 4 are in Table 2.

TABLE 2  
MEAN PER CENT ERRORS AS A FUNCTION OF FREQUENCY CLASS FOR THE TWO GROUPS  
AT BOTH ASLs ON DAY 1 AND DAY 4

		Class of item frequency											
		I				II				III			
		HR		NR		HR		NR		HR		NR	
ASL	Day	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
3.5	1	47	13	50	16	42	18	49	10	47	14	51	10
	4	19	12	37	20	17	10	41	17	29	15	32	15
6.5	1	64	13	78	9	72	10	73	12	70	11	75	7
	4	40	15	63	11	55	13	68	13	64	12	65	11

Analyses of variance between the groups were computed for the three item classes at both ASLs. On Day 1 Group HR made fewer errors than Group NR for Class I items at the 6.5 ASL ( $F = 7.86$ ;  $df = 1,18$ ;  $p < .05$ ). The groups did not differ significantly at any of the other five comparisons. On Day 4 Group HR made fewer errors than Group NR for Class I items at 3.5 ASL ( $F = 5.21$ ;  $df = 1,18$ ;  $p < .05$ ) and at 6.5 ASL ( $F = 12.72$ ;  $df = 1,18$ ;  $p < .01$ )—and for Class II items at 3.5 ASL ( $F = 13.20$ ;  $df = 1,18$ ;  $p < .01$ ) and at 6.5 ASL ( $F = 4.62$ ;  $df = 1,18$ ;  $p < .05$ ). The groups did not differ in their recall of Class III items.

#### D. DISCUSSION

The results reported here have been replicated with a slightly different design. In the replication, four groups of Ss which differed systematically in the range of item frequencies were tested. The frequency ranges for the four groups were 64-1, 48-1, 32-4, and 16-4. Total errors increased systematically from Group 64 to Group 16. Per cent errors for Class I items also increased as range decreased. Again, the higher-range groups either recalled more Class II items or at least as many as the lower-range groups.

The purpose of the studies was to learn the effects on recall of varying the frequency of occurrence of the items presented for storage. The overall effect was to improve total recall. The facilitation due to varying frequency was not limited to high-frequency items (Class I), but persisted for low-frequency items (Class II). These results indicated that HR Ss were not remembering high-frequency items at the expense of low-frequency items.

The data do not directly provide an explanation of how frequency variations facilitated recall. A simple hypothesis that Ss learn to match response probabilities with stimulus probabilities does not seem adequate because

Group HR Ss did well on both Class II and III items. One possible explanation might be that the HR Ss, like the E, acquired some means of discriminating groups of items. They could have grouped or chunked (4) the items into high- and low-frequency classes, for example, and thus eased their storage load. As range decreased it became more difficult for Ss to discriminate the item classes, and hence more difficult to chunk the items on that basis.

The purpose of manipulating ASL was to test for an  $ASL \times Range$  interaction. This interaction was not significant. Lloyd and Johnston (2) did not obtain an analogous interaction between ASL and contextual constraint. It may be that this interaction will occur only when extremely small ASLs (e.g., 1 or 2) are employed.

### E. SUMMARY

The daily frequency of occurrence of items (letter-word pairs) was varied in a sequential memory task. In this task, S continually received information and, at unpredictable moments, was requested to recall some of it. Two groups of Ss which differed in the range of variation in item frequencies in the sequences presented to them were compared. Total recall was directly related to range of item frequency ( $p < .01$ ). High-frequency items in a high-range group were recalled more often than medium-frequency items in a narrow-range group ( $p < .01$ ). Items with a low frequency in both the groups were recalled most often by the highest-range group ( $p < .05$ ). The facilitation of recall in the highest-range groups was not limited to the high-frequency items.

### REFERENCES

1. HALL, J. F. Learning as a function of word frequency, *Amer. J. Psychol.*, 1954, **67**, 138-140.
2. LLOYD, K. E., & JOHNSTON, W. A. Short-term retention as a function of contextual constraint. *J. Exper. Psychol.*, 1963, **65**, 460-467.
3. LLOYD, K. E., REID, L. S. & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. Exper. Psychol.*, 1960, **60**, 201-207.
4. MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81-97.
5. SUMBY, W. H. Word frequency and serial position effects. *J. Verb. Learn. & Verb. Behav.*, 1963, **1**, 443-450.

Department of Psychology  
Washington State University  
Pullman, Washington

## PARTICIPATION, INTERACTION AND ROLE BEHAVIOR IN GROUPS SELECTED FROM THE EXTREMES OF THE OPEN-CLOSED COGNITIVE CONTINUUM\*

*Department of Psychology, University of Arizona*

---

SALVATORE V. ZAGONA AND LOUIS A. ZURCHER, JR.

---

### A. INTRODUCTION

A great deal of research time has been spent in describing the personality structure and dynamics of the individual who has been variously characterized as "authoritarian," "dogmatic," or "rigid." The authoritarian personality, for example, is described as adhering rigidly to conventional middle-class values; as having a submissive, uncritical attitude toward idealized moral authorities of the ingroup; as tending to be on the lookout for, and to condemn, reject, and punish people who violate conventional values; as being opposed to the subjective, the imaginative and the tender-minded; as believing in mystical determinants of the individual's fate and as having the disposition to think in rigid categories; as being preoccupied with the dominance-submission, strong-weak, leader-follower dimension; as having generalized hostility; and as being disposed to believe that wild and dangerous things go on in the world, thus projecting outward unconscious emotional impulses [Adorno *et al.* (1, pp. 248-250)].

The theory of dogmatism was formulated by Rokeach as an alternative to the concept of authoritarianism because he felt that authoritarianism, particularly as measured by the F Scale, is a measure of *right* authoritarianism and *right* intolerance; whereas, dogmatism, as measured by the Dogmatism Scale, is a measure not only of *general* authoritarianism and *general* intolerance, but also of the "openness-closedness of cognitive systems." In a series of papers (3, 4, 5) and in a book (6), Rokeach describes the empirical studies and the theoretical logic that led him in the formulation of his theory.

The dogmatic individual is best described by considering Rokeach's definition of the concept of dogmatism: "a relatively closed cognitive organization of beliefs and disbeliefs about reality organized around a central set of beliefs about absolute authority which, in turn, provide a framework for patterns of intolerance and qualified tolerance toward others" (5, p. 3). Hence, the

---

\* Received in the Editorial Office on May 8, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.



dogmatic individual can be said to be one who (*a*) accentuates the differences between his belief and disbelief systems, (*b*) denies evidence that is contrary to his belief systems, (*c*) allows contradictions to exist within his own belief systems, and (*d*) disregards as irrelevant similarities between his belief and disbelief systems (5, pp. 7-9).

Authoritarian and dogmatic individuals have been observed in a wide variety of experimental situations, and the relationships between "F," "E," or "D" and an enormous number of other personality and situational variables have been (since 1950) the subject of many hundreds of studies—theses, dissertations, and journal articles.

There seems to have been a shift of research emphasis away from the purposes that led to the original identification of "The Antidemocratic Personality" (2). That research stemmed from a concern with identifying individual personality correlates of *intergroup* tensions. While this point of view is implicit in the large number of studies following the publication of "The Authoritarian Personality" (1), the emphasis has gradually shifted from a concern with the effects (upon the *group*) of an authoritarian outlook to a concern with the manner in which authoritarianism (or dogmatism) in an individual is related to other characteristics that the individual may possess.

## B. PURPOSE

The purpose of the present study is to determine the extent to which dogmatic persons and nondogmatic persons will demonstrate their characteristics in group situations or to determine the manner in which group processes will reflect the attributes of the memberships of the groups involved.

We are not so concerned with describing the low-dogmatic or high-dogmatic individual as we are with observing the behavior of such persons when each is in interaction with others of his own kind.

Inherent in all research on attitudes is the problem of determining the extent to which scores on attitude questionnaires are predictors of the implied behavior. Put in another way, the question concerns the degree of correspondence between elicited verbal attitudes and action attitudes. Do subjects who score high on scales, such as the E scale, the F scale, or the Dogmatism Scale, do so because of naiveté and, therefore, show themselves in an unfavorable light in virtually any personality measure applied to them? Or do the scores of such persons reflect the degree to which they *behave* in an ethnocentric, authoritarian, or dogmatic manner? The persistent high intercorrelations among the measures mentioned causes one to suspect that some personal variable—a response tendency, perhaps—that is only indirectly related (if

at all related) to the variable in question may be operating in each of the test situations. Rokeach (6, pp. 405-406) raises question as to whether or not response set might be a contributing factor in high-dogmatism scores and whether or not his many research findings can be accounted for by response set. The possibility that docility, or submissiveness, may contribute to a passive acceptance of Dogmatism-Scale items ("agree" responses contribute to a high-dogmatism score) is an interesting one. Submissiveness to authority is a characteristic of the highly dogmatic individual; if we were to observe such a trait in our High-Dogmatic group (which we did, to a significant degree), it would not be a particularly meaningful bit of evidence *in itself* for, paradoxically, this evidence can be used as an argument to reject the validity of the Dogmatism Scale as well as to support it.

But other behavioral traits related to the theory of dogmatism should clearly distinguish our two groups. How would each individual react in group-decision situations of an ambiguous nature? Would communication channels in group problem-solving tasks be relatively free or restricted in one group as compared with communication channels in the other? In what ways would the groups differ with respect to the manner in which leaders are chosen, in their reactions to challenges to authority figures, and in their ability to arrive at group consensus in discussions of a controversial and personal nature?

It is not often that elicited verbal attitudes, such as those tapped by the Dogmatism Scale, can be subjected to a systematic observational procedure that enables their behavioral correlates to be observed over a relatively long period of time and under a variety of circumstances. If it is demonstrated that distinctions made on the basis of verbal responses are confirmed by behavior in specially created social situations, then significant evidence of the validity of the Dogmatism Scale is obtained.

### C. HYPOTHESES

In the present investigation, the following hypotheses, derived from Rokeach's formulations (3, 4, 5, 6), were tested.

#### 1. *In the Classroom*

The High-Dogmatic group is *leader-oriented* and its members prefer lectures to discussion. If held, novel points of view are rarely expressed and spontaneity is minimal.

The High-Dogmatic group shows a preference for clearly structured topics and instructional situations. Its members are typically uncreative and they are routine and conventional. They are disturbed by behavior (on the part

of the instructor) that does not conform to their expectations of the role behavior of an authority figure.

### *2. Ingroup Interaction Experiments Dealing with Controversial Material*

1. There is a more active concern (preoccupation, perhaps) on the part of High Dogmatics with the problem of leader selection. Rokeach's theory suggests that the dogmatic individual, whether or not he identifies with the leader or the follower, is at any rate leader-follower oriented.

2. As in the classroom situation, the need for structure overshadows the need for expression of spontaneity.

3. When group consensus is reached, challenges by authority figures produce reactions that are characteristic of the High- and the Low-Dogmatic groups: the High-Dogmatic group becomes insecure, wavers in its conviction, and evidences signs of reduction in group cohesion. On the other hand, the Low-Dogmatic group tends to defend its consensus and, if anything, becomes more unified by a challenge from the authority figure.

4. Group consensus by the Low-Dogmatic group is reached with more difficulty: i.e., after more time is spent in discussion, than is the case in the High-Dogmatic group; however, when the two groups are brought together to present and to defend their decisions and to arrive at a general consensus, the Low-Dogmatic group-decision prevails.

### D. SUBJECTS AND METHOD

Rokeach's Dogmatism Scale was administered to 517 undergraduate students enrolled in the elementary psychology course at the University of Arizona. From among the total *N*, the 30 highest and the 30 lowest scorers were assigned to two special conference sections—the High Dogmatics to one section and the Low Dogmatics to another. (Students in the elementary psychology course attend two general lectures and one conference section weekly. Each conference section contains approximately 30 students; thus the experimental groups were similar to other conference sections.)

The authors, as instructors for the conference sections, utilized the classroom to make semester-long observations of the characteristic performance and nature of interaction of the High- and the Low-Dogmatic sections respectively. In addition, subjects were drawn from the two conference sections for small-group studies. The purpose was to determine systematic differences between High- and Low-Dogmatic groups in level of interaction, in facility in determining group consensus on controversial issues, in development of leadership, in responses to challenges by authority figures, and in any other

pertinent distinguishing characteristics that might appear during experimental observation.

Two experimental rooms, furnished in seminar fashion (and complete with one-way mirrors, concealed microphones, and inputs to tape recorders) were utilized for the experimental observations. The rooms are constructed so that both can be surveyed from a central observer's room. Tape recordings of the proceedings were made and, in addition to the two experimenters, 10 other observers (graduate and undergraduate students in the experimental social psychology class) systematically noted the occurrence of data relevant to the experimental variables.

In study No. 1, six subjects drawn from the High-Dogmatic conference section were assigned to one observation room, and six subjects from the Low-Dogmatic conference section were assigned to the other. The 10 observers were not informed which group was which.

To both the High-Dogmatic and the Low-Dogmatic groups a standardized statement was read that presented the following controversial question for discussion:

Some people think that there is an absolute truth, or a source of absolute truth, upon which thoughts and deeds can be unerringly interpreted as being either right or wrong.  
What is your opinion?

Each group was asked to come to a general agreement in 20 minutes. Then one of the group would report the consensus to the experimenter. The experimenters then left the rooms and joined the other observers. After a group reached a general agreement that its members thought could be presented, one of the experimenters entered the room, listened to the report, and immediately took issue with it. He challenged the decision, ridiculed the group, and argued with its members in an attempt to get the group to change its view to his: a view that, by plan, was opposite to that expressed by the group.

Study No. 2 was conducted in the same manner as Study No. 1. Twelve subjects (six High-Dogmatic subjects and six Low-Dogmatic subjects) were told that, on a controversial question (different from that used in Study No. 1), they would present their conclusions to another group and, after an intergroup discussion, would come to a "grand consensus."

To both groups, the following statement was read:

The question of men being permitted to entertain women in college dormitories and vice versa has been widely discussed recently, particularly with reference to the "goings on" at large Eastern universities. While some psychiatrists may believe that premarital sexual relations



may be beneficial as a form of tension release, this view is not universally accepted.

The topic for your discussion is, therefore: There is nothing wrong with having an affair with someone you love, but do not expect to marry. In fact, such affairs could only benefit, by teaching you sexual technique, both you and the person you will later marry.

After they reached their respective decisions, the High-Dogmatic and the Low-Dogmatic groups were brought together in one observation room and were given time to report to each other and to reach a second conclusion. Then they were challenged by the experimenter.

### E. RESULTS AND DISCUSSION

It should be noted that the basis of comparison was more than a mere comparison of one group with the other,—it was how each group's performance compared with those of the many other groups taught by the writers in non-experimental situations. It is acknowledged that there are wide differences in spontaneity, alertness, extent of classroom participation, etc., among relatively small lecture-discussion groups. However, the writers were unprepared for the extent of the differences that were immediately observed between the High- and the Low-Dogmatic groups.

Our first hypothesis predicted that the High-Dogmatic group would be leader-oriented, unspontaneous, and would prefer lectures to class discussion. The data relevant to our hypotheses are shown in Tables 1 and 2.

It can be seen that virtually every hypothesis suggested by the theory of dogmatism and tested in this investigation has been confirmed. Perhaps for the first time since research has been conducted on Rokeach's dogmatism theory, groups of high-dogmatic and low-dogmatic individuals, respectively, have been experimentally formed. These groups remained intact as social units for over four months and were observed in a wide variety of situations demonstrating intragroup and intergroup dynamics. Comparisons between the groups on leader orientation, spontaneity and originality, attitudes toward authority, and group cohesiveness in the face of challenges from authority figures show important differences and confirm significant aspects of dogmatism theory.

In general, high-dogmatic individuals as a group and low-dogmatic individuals as a group conformed to the behavior patterns one would expect in situations contrived in accordance with the theory of dogmatism. There were, however (especially in the experimental discussion groups), occasions when a subject temporarily seemed to adopt a role contrary to his cognitive



TABLE 1  
HYPOTHESIS A—CLASSROOM INTERACTION: SPONTANEITY AND ORIENTATION  
TOWARD LEADERSHIP

Low-Dogmatic group	High-Dogmatic group
Given the opportunity, preferred to be graded on basis of preparation of original and creative individual research effort.	Rejected opportunity to be graded in terms of individual research; preferred to be graded in terms of objective well-structured examinations.
Spontaneously engaged in animated discussion; extended discussion to entire group frequently, with remarks and rejoinders.	Preferred lecture-type presentation; appeared uncomfortable unless material was presented in outline form, for easy note-taking.
Concerned mainly with psychological issues and how these pertain to general human problem areas.	Concerned mainly with learning the facts of psychology upon which they would later be tested.
Leader's role in the group often relegated to that of "catalyst" of discussion. Discussion often bounced back and forth from class member to class member and became so heated that the instructor was virtually ignored for periods of time.	Instructor was obviously perceived as the leader of the group; class depended upon him for leadership in every phase of classroom activity.
Preferred the final examination in the course to be of the essay type.	Unanimously expressed a desire for a multiple-choice final examination.
Not overly concerned with complying with minor regulations. Not embarrassed by tardiness, frequently continued class discussion long after the bell had sounded. Spirited discussions frequently were observed to continue in the hallways and outside the building.	Very seldom tardy for class; always prepared to leave the classroom the moment the bell sounded, regardless whether a task or discussion had been completed.
Appeared to be relaxed, though dissatisfied with the present state of political, social and economic conditions. Freely expressed annoyance with arbitrary authority figures and bureaucratic academic procedures.	Tension and anxiety were the most conspicuous characteristics of this group. Concern with conforming to academic regulations and expectations.

framework as indicated by the Dogmatism Scale. This was especially true when a subject was under pressure from the experimenter or from a fellow group member. High Dogmatics were observed to make such statements as, "We have to look at both sides of the question" and "Let's be broad-minded about this." Low Dogmatics were observed to say, "There's no arguing this point—I can't be convinced otherwise," and, "If the Church says so, that's good enough for me."

It would appear, therefore, that although long observation of a group of High and Low Dogmatics reveals a relatively consistent behavioral pattern, there are (in given social situations) times when individuals adopt role be-

TABLE 2  
HYPOTHESIS B—SMALL-GROUP STUDIES: DISCUSSION OF CONTROVERSIAL MATERIAL

Low-Dogmatic group	High-Dogmatic group
<p>Selection of a committee chairman involved little or no discussion. It appeared not to matter who emerged as chairman; instead, the concern was with promoting the discussion, and with exchanging ideas.</p> <p>The most casual methods were employed in the selection of the chairman. Remarks typically made in this connection: "You're closest to the pencil and paper; how about you keeping the records?" The group apparently did not perceive the chairman's role as one of leadership; rather, that of simply recording the progress of discussion. All members joined in the discussion so that it was virtually impossible to note that a leader had, in fact, been chosen.</p> <p>When reports were made of the final group decision, the chairman had difficulty in making the report without repeated corrections, additions, and other interruptions from various members of the group.</p> <p>Immediately after the controversial statement had been presented to the group for discussion, they displayed an impatience to begin exchanging ideas. The group appeared to be so preoccupied with the challenge offered by the subject that they hardly noticed the departure of <i>E</i>.</p> <p>In an apparently relaxed and straightforward manner, the group members began to air their views. The subject of sex was treated openly and freely, with little embarrassment in the use of direct terminology for sexual behavior. There seemed to be no apprehension about becoming personally identified with sexual intimacy. Personal pronouns ("I," "you," etc.) were used freely in the discussion. ("I would sleep with a man only if," etc. Or, "Would you have sexual relationship with . . ." etc.)</p>	<p>Embarrassed clumsiness in the selection of a chairman and seeming unwillingness by the member to assume the leader role. Members seemed to attach a great deal of importance to the question as to who should be the chairman; some felt that the experimenters should have appointed a "leader." Once decided, the chairman guided and structured the discussion, in addition to recording the proceedings. Reports of the group decision were made solely by the chairman, without challenge or correction by any of the group members. At the conclusion of the report by the chairman, the group was asked by <i>E</i> whether or not the reported consensus accurately reflected views held by the group members. The group submitted in silence to the report rendered by the chairman, even though elements of the report had been noted by the observers to be in direct conflict with the previously expressed views of various group members.</p> <p>This group appeared to be leader-oriented. They were clearly reluctant to have <i>E</i> leave them with an unstructured problem and an unstructured group situation. They asked for more instructions. Questions appeared that indicated a need to have the whole situation more clearly formalized: "How do you want us to pick a leader?" "How do you want us to come to an agreement?" "What kind of answers are you looking for?" "What form do you want us to make this report in?"</p> <p>After the departure of <i>E</i>, the group fell into an awkward silence, followed by nervous laughter, and "what do we do now?" Tension and anxiety seemed to prevail, particularly before and during the discussion on premarital sexuality.</p> <p>The matter of sex was treated in a detached, impersonal manner, and, even so, with considerable embarrassment. Group members discussed the subject as though sex were a behavior performed by some alien, remote race, and not something that they themselves might conceivably some day participate in. Phrases used: "doing it," "they may get together if they love each other," or—referring to the sexual act—"committing sin."</p>

## Low-Dogmatic group

The group met *E*'s challenge to their consensus report with a consistently unified and logically argumentative defense. The members were willing to question the argument of the authority figure *vis-à-vis* the veracity and tenability of their own position. The statements by *E* were interpreted as opinion, really no more nor less correct than their own opinions.

Because of greater member participation and a higher level of interaction, a significantly longer time was taken to reach consensus. When merged with the High-Dogmatic group, the Low-Dogmatic group members virtually dominated the discussion, the comments of the High-Dogmatic group members being almost limited to the report of their consensus. When a "grand" consensus was required to be reported (i.e., a consensus representing the views of both groups), this merged group consensus was drafted and reported by the Low-Dogmatic group. Finally, when *this* report was challenged by *E*, the Low-Dogmatic group provided the main substance of defense, the High Dogmatics playing relatively little part.

## High-Dogmatic group

The group met the challenge of their consensus report with, at first, some little resistance. As *E* persisted, however, the fabric of group cohesion showed marked signs of weakening; individual members began questioning each other as to who was responsible for those parts of the consensus now under fire by *E*. Whatever group solidarity could be said to have existed prior to the threats posed by *E* was now entirely disintegrated. In a relatively short period of time, members assumed the role of passive listeners to *E*'s arbitrary comments. Subsequently, when *E* asked the members if they now saw the errors in their decision, they all to a man agreed.

This group took a relatively short time to come to an agreement, but seemed anxious about joining with and reporting to the other group. They seemed to have little confidence in the "presentability" of their consensus and expressed concern that they would be "shot down" by the others. Following the introduction of the two groups to each other, the chairman of this group apologetically reported its consensus. From that point on, he and his fellow members had little to say, quietly acquiescing to the structure provided by the Low-Dogmatic members.

havior that is not in line with their position on the continuum of the dogmatism variable. If an appraisal of an individual, in terms of high or low dogmatism, was made on the basis of his behavior in an isolated moment, the observer might classify the individual in precisely the wrong category.

Rokeach noted the impact of the social situation when he found that professors cannot accurately judge, even on the grossest level, the dogmatism of graduate students they know personally. Apparently, graduate students can "mask" the closedness of their belief systems. As Rokeach points out, "In such a situation, he [the graduate student] is likely to adopt a respectful, acquiescent, and, at the same time, an enlightened and objective facade" (6, p. 106). Such behavior seemed to be true of our subjects, at times, especially when they were unduly pressured by experimental authority figures or by

group demands. In essence, it seems that, on the behavioral level, the momentary expectations of a social role can override personality constellations measured by attitude questionnaires.

#### REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The Authoritarian Personality*. New York: Harper, 1950.
2. FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The antidemocratic personality. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in Social Psychology*. New York: Holt, 1947.
3. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, **61**, 194-204.
4. ———. On the unity of thought and belief. *J. Personal.*, 1956, **25**, 224-250.
5. ———. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monog.*, 1956, **70**, No. 18.
6. ———. *The Open and Closed Mind: Investigations into the Nature of Belief Systems and Personality Systems*. New York: Basic Books, 1960.

*Department of Psychology*  
*University of Arizona*  
*Tucson, Arizona 85721*

## CHEMICAL STUDIES OF BEHAVIOR: VIII. SOME ANOMALOUS ONSET EFFECTS OF BENZQUINAMIDE, A "TRANQUILIZER"\*

*Department of Psychology, University of Rochester*

---

JEAN S. CAMERON AND G. R. WENDT<sup>1</sup>

---

### A. INTRODUCTION

The new drug, benzquinamide,<sup>2</sup> was submitted to us for evaluation of its effects on moods, emotions, and motivations in normal young men. Extensive clinical and experimental investigations had indicated that it had a tranquilizing action (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11). We agreed to incorporate tests of the drug into a large experiment in which a number of other drugs were included (placebo, amphetamine sulphate, dimenhydrinate,<sup>3</sup> and others).

While our data from doing this experiment showed that benzquinamide was indeed a sedative drug when the effects of a single dose were assessed two hours after administration, we were sharply surprised by the presence of earlier onset effects that were stimulating to a number of emotional systems and alarming to the subjects. The purpose of the present report is not only to describe these anomalous onset effects of a "tranquilizer," as indicated in our title, but also to raise an important question (Why did onset effects escape the attention of clinical investigators working with patients?), and to report briefly on methods of avoiding the onset effect by divided doses or sustained release medication and on the nature of the steady-state effects of the drug.

### B. METHODS

The same methods have been used by us—in experiments done in 1954, 1955, 1958, 1959, 1962, and 1963—using a variety of drugs. The procedures are designed to make an assessment of drug-induced changes in a wide variety of moods, emotions, and motivations and to make assessment in

---

\* Received in the Editorial Office on May 11, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> We acknowledge with gratitude the collaboration of Priscilla G. Specht, who shared the responsibility of conducting the experiments.

<sup>2</sup> Trade-named Quantril; made by Chas. Pfizer and Co., Inc. This work was done in the spring of 1962, except for some follow-up phases done in June and December of 1962. It was partly supported by a grant from the Pfizer Company. Our scientific liaison was through Dr. James L. Garofalo and Dr. Leonard S. Brahen.

<sup>3</sup> Dramamine, made by G. D. Searle and Co.



such a way that the "placebo effect" is eliminated and a high degree of sensitivity is achieved (12, 13). The main data obtained are from Adjective Check Lists (12; p. 205, 207). These ACLs give the subject an opportunity to indicate which of a list of self-descriptive adjectives apply to his feelings at that time (e.g., cooperative, good-natured, cheerful, decisive, forceful, impulsive, active, businesslike, easygoing, quiet, sleepy, tired, belligerent, impatient, apprehensive, skeptical, disturbed, confused, blue, depressed, subdued, withdrawn). One ACL contains 132 such adjectives; another contains 34 pairs of adjectives (e.g., active-energetic, gloomy-blue).

In the work described here, 50 male college students, aged 21 or slightly older, served in a number of sessions, the sessions being separated by two or more days. On each day the *S* came twice: once in the late morning, when he filled out report forms, completed an ACL, and took his medication (in No. 1 clear gelatin capsules); again, two hours after medication (one and one-half hours in some cases), to report on the intervening two hours and to complete ACLs [see Wendt and Cameron (12) for details]. A number of drugs and placebos were included in the study, different *Ss* being on different drug sequences to counterbalance variable factors. (Seven of the 50 *Ss* served at two-dose levels of benzquinamide.) *Ss* came to the laboratory individually. The experiment was double-blind except that at the conclusion of each session, after all the data of an *S* were collected, the experimenter looked up the nature of the medication as a precautionary measure followed in the use of experimental drugs. The benzquinamide sessions were imbedded in three simultaneous drug experiments using 73 *Ss* in 620 sessions.

Our previous work has shown a satisfactory degree of repeatability of data when drugs such as Dramamine and Benzedrine are used (13). By calculating the drug-induced change in each adjective in each experiment and then correlating the 132 pairs of adjective change scores from any two experiments, we found, for example, that after 80 mg of Dramamine the data of separate experiments are highly correlated ( $r = .84$  to  $r = .92$ ).<sup>4</sup> The absence of placebo effects is shown by  $r$ 's ranging from  $-.07$  to  $.05$  between experiments, showing an absence of systematic factors (13).

### C. DATA ON ONSET EFFECTS

Based on clinical data and on tests on ourselves (unfortunately done by taking several divided doses separated by 30 minutes),<sup>5</sup> we concluded that 150

<sup>4</sup> Unpublished data obtained on the short ACL show correlations as high as .98 for repetition of experiments on Dramamine. We routinely control all of our experiments not only with placebo but also with Dramamine and Benzedrine.

<sup>5</sup> Or perhaps, fortunately, since it led to the unintended study of onset effects.

mg of benzquinamide would be needed to produce a readily detectable effect. But as soon as we began to give that amount in a single dose we ran into trouble because we were dealing with college students who were expecting no strong medication. Every one of the first five Ss was seriously disturbed and went to bed on his own initiative. We then reduced the dose to 75 mg (13 Ss) and continued to have trouble; then, to 50 mg (three Ss) and again had trouble; finally, to 25 mg (26 Ss), when only three Ss reported anomalous onset effects. Altogether, there were 19 instances (in 57 drug administrations) in which S either volunteered information concerning the presence of sudden onset effects or gave it in response to questioning about the effects of the drug. We then selected three Ss who had had strong onset effects and studied the Ss more intensively under continuous observation in the laboratory. In these Ss the onset effects lasted from 10 to 30 minutes, and the transition at the end of that period was brief (five minutes). One graduate student, not in our original group, was studied after 75 mg and showed a very complete picture of onset effects.

We collated the descriptive terms volunteered by the Ss who felt anomalous onset effects. The following lists include most of them (the words have been rearranged to bring together those of similar meaning):

1. restless, tense, nervous, jittery, keyed-up, excited.
2. uneasy, apprehensive, anxious, trapped.
3. confused, impossible to describe.
4. unreal, remote, shut-in, lost, completely closed-in.
5. irritable, impatient, antagonistic, annoyed, hostile, angry, resentful, rebellious, suspicious, unfriendly, resistant, uncooperative.
6. unsociable, withdrawn, uncommunicative, self-contained, close-mouthed.
7. depressed, unhappy, bad, disturbed.
8. unmotivated, unenthusiastic.
9. weak, tired, sleepy, lazy, dull, lethargic, blah, pfft.
10. dizzy, light-headed.
11. nausea, loss of appetite.
12. headache.

Other descriptive words in the same general classes were checked or double-checked on the required Adjective Check Lists, but these were (at two hours after drug) a blend of words expected after a sedative (e.g., easygoing, indifferent, leisurely) and words reflecting the carry-over of strong negative feelings about the onset effect.

Comments included:

It felt as though I were standing back looking at myself. It felt unreal (150 mg).

It's such a helpless feeling. I want to run away from it but there isn't any place to run to (75 mg).

I feel completely uncooperative. . . . the feeling is almost impossible to contend with. There's nothing you can do about it (50 mg).

One *S* developed a marked depression of 15 minutes' duration. He reported himself as severely emotionally depressed, remorseful, guilty and "completely closed in." Others reported depressions of lesser degree. It may also be noted that one *S* (150 mg) developed severe hiccups (duration one hour), another *S* (50 mg) developed slight hiccups, one *S* complained of warm skin, and one *S* developed forced laughing (laughter was his characteristic method for handling depression).

#### D. STEADY-STATE EFFECTS OF BENZQUINAMIDE

Definitive statements about the steady-state effects of the 150-mg dose are hampered by the contamination of the data by a carry-over of resentment and apprehension resulting from the unexpected onset effects. The least seriously contaminated data are those from the 13 *Ss* who received 75 mg of drug. We "corrected" the frequency of checking of each adjective after drug by the average of the responses on placebo days (at least three days for each subject) and determined those adjectives with the largest corrected change scores (13, pp. 260-266) after drug. In rank order of amount of drug effect, the adjectives showing the greatest *increases* in checking after 75 mg of benzquinamide were:

sluggish, light-headed, sleepy, drowsy, lazy, leisurely, tired, weary, lackadaisical, subdued, washed-out, dizzy, keyed-up, weak, apprehensive, dull, quiet, depressed, disturbed, relaxed, close-mouthed, emotional, and jittery.

Those showing the greatest *decreases* in checking (in rank order) were:

attentive, ambitious, self-confident, efficient, eager, industrious, enterprising, resourceful, work-oriented, task-involved, energetic, active, sociable, businesslike, masterful, hardworking, confident, cautious, assertive, willful, patient, and engrossed.

Most of these adjectives are among those similarly altered by tranquilizers. Adjectives such as apprehensive, disturbed, and jittery are looked upon as carry-overs of the onset effect.

The data after 50 mg were too few (three *Ss*) to be worth reporting. The 25-mg data (26 *Ss*) show that this dosage on the average was ineffective: only one of 132 adjectives (hardworking) increased enough barely to reach

our criterion (13, p. 264) for inclusion as a probable drug effect. The adjectives within one and two checks of reaching our criterion were (*increased*) weak, elated, eager, lackadaisical, tired, subdued, lazy, industrious, meticulous, washed-out; (*decreased*) effective. Such mingling of positive and negative adjectives can be produced by subgroups of Ss who change in opposite directions. (There is some evidence that a very small dose of benzquinamide produces a positive emotional state.)

The magnitude of the steady-state effect obtained after 75 mg of benzquinamide is about the same as we previously obtained from 50 mg of Dramamine, less than that obtained after 80 mg of Dramamine (13), and greater than that from 600 mg of meprobamate (four unpublished experiments).

The effects of three dose levels of benzquinamide (150, 75, and 25 mg) on subjects' choices on the short, forced-choice ACL (12, p. 207) are displayed in Table 1. (In this ACL the S is required, for each adjective pair, to indicate whether he felt more that way "before" drug, or feels more that way "now": i.e., two hours after drug.) Minus numbers in the table indicate that a smaller percentage of Ss reported feeling more that way after drug than after placebo. Plus numbers indicate a greater percentage feeling more that way after drug. The data from the 50-mg dose are omitted because there were only three Ss (the results were similar to those after 75 mg). For ease of comprehension, the sequence of words in the table has been rearranged from that in the ACL.

From Table 1 it can be seen that the effects of 150 mg and 75 mg of benzquinamide are similar, each of the 34 items of the ACL being changed in the same direction for both groups of Ss. In 30 out of 34 items the 150-mg dose affected a greater percentage of Ss than did the 75-mg dose, but this may be partly an artifact of the placebo data calculated from only five Ss. The changes due to benzquinamide are (on the whole) like those due to Dramamine or meprobamate, but differ in that there was more evidence of negative emotions: e.g., jittery-nervous, anxious-fearful, disturbed-upset, grouchy-irritable, impatient-snappish. We interpret these as carry-overs from the onset effects.

The results obtained from 26 Ss after 25 mg of benzquinamide showed the effects to be small and apparently mostly in a stimulating direction, rather than sedative. In 23 out of 34 items, the apparent effects were opposite to those after larger doses.

Item 34 in Table 1 shows the drug-placebo comparison on the question, "Which way of feeling did you like better? Before? Now?" The larger doses



TABLE 1  
EFFECTS OF BENZQUINAMIDE ON FORCED-CHOICE ACL  
(DIFFERENCES BETWEEN DRUG AND PLACEBO)  
(Percentage of subjects)

Adjective pair	150 mg (N = 5)	75 mg (N = 13)	25 mg (N = 26)
1. active—energetic	-47%	-36%	+ 5%
2. effective—efficient	-60	-41	- 5
3. industrious—work-oriented	-47	-41	+ 7
4. decisive—capable	-60	-38	0
5. self-confident—courageous	-53	-33	- 5
6. egotistic—boastful	-47	-26	-12
7. talkative—chatty	-60	-16	+13
8. humorous—witty	-60	-18	+18
9. genial—friendly	-67	-29	+10
10. cheerful—happy	-67	-46	+ 9
11. intoxicated—light-headed	+33	+11	- 8
12. easygoing—relaxed	-27	-10	+ 4
13. quiet—peaceful	-40	- 2	-11
14. careful—meticulous	-33	-31	+ 8
15. self-conscious—timid	+20	+33	0
16. jittery—nervous	+40	+26	+10
17. anxious—fearful	+60	+39	-11
18. guilty—remorseful	+27	+15	- 7
19. downhearted—sad	+73	+18	- 1
20. gloomy—blue	+53	+18	- 3
21. pain—discomfort	+60	+18	+ 1
22. disturbed—upset	+47	+41	+ 2
23. confused—disorganized	+80	+36	+ 2
24. drowsy—sleepy	+40	+33	0
25. tired—washed-out	+27	+41	+ 6
26. weak—lacking strength	+33	+39	+ 9
27. dull—sluggish	+40	+20	+ 4
28. uninterested—bored	+13	+15	+ 2
29. resentful—indignant	+33	+ 2	-10
30. angry—annoyed	+40	+ 5	- 1
31. grouchy—irritable	+73	+38	- 7
32. defiant—belligerent	+13	+ 3	-13
33. impatient—snappish	+27	+20	- 2
34. Which way did you like better?	-73	-39	0

were disliked, as are the effects of sedatives such as Dramamine and meproba-  
mate. Under the conditions of these experiments, our Ss, young healthy  
college students, do not regard sedation as a happy or positive state. (If on  
Item 34, we combine all answers obtained on placebo days in five different  
experiments, we find that 52 per cent of Ss felt better after lunch.)

#### E. METHODS OF AVOIDING THE ONSET EFFECTS

We hypothesized that one of the reasons clinical investigators had failed  
to report the onset effects was that drug administration three or four times



each day resulted in a residual of the drug that overcame the onset. We did two brief experiments to test this hypothesis, and the obtained data were in accord with it.

In the first experiment, observations were made on three of our former Ss, each of whom had previously had a strong onset effect. Each S took three doses of 50 mg on one day and two on the next. One dose of each day was administered before breakfast and another before lunch. The S stayed under continuous social interaction with the experimenters until the onset effects were over. (This work was not double-blind; the Ss were told the problem.) Onset effects were obtained on each of the 12 occasions when the Ss were under observation, but differed greatly in degree. The strongest onset effects were in reaction to the first dose of the first day; the next strongest effects were in reaction to the first dose of the second day; the smallest effects were in reaction to the second administrations of each day, which administrations merely produced some restlessness and social withdrawal. Because we did not have the Ss under observation after the third administration of the first day, we feel that their reports of lack of onset effects cannot be confidently accepted.

In a second experiment, we used 150 mg of drug in a sustained-release form designed to release about 23 mg per hour. Seven normal Ss with previous experience in our work had a total of 16 administrations. (Two of these Ss had shown strong onset effects from earlier doses; the remainder had no experience with larger doses of benzquinamide.) No subject was able to detect anomalous onset effects from the sustained-release medication. There were indications in the ACLs (obtained at two hours and six hours after drug) that suggest further research on the time-course of action. The two-hour reports were of positive emotions (e.g., cheerful, friendly, relaxed) reminiscent of the short ACL data after 25 mg, but in the six-hour reports about half of the Ss reported a reversal to some restlessness and irritability. These ACL results are suggestive, not dependable, because they were obtained under less than ideal conditions.

## F. DISCUSSION

### 1. *Why Did Some Clinical Investigators Overlook the Onset Effects?*

At the time we undertook this work a number of clinical investigators had administered benzquinamide to a large number of patients (ca.1,000). Yet the makers of the drug had no knowledge of onset effects. Since then a considerable number of investigators have found anomalous effects, (e.g., 1). We suggest that some of the reasons for failure to find them may be as follows:

1. Clinical investigators typically do not see a patient until several days after starting t.i.d. administration. By that time the patient may be having no onset effects and thus fail to report on the first dose.

2. Patients may mistake the onset effects for symptoms of their emotional illness and thus not report them.

3. Often patients try to please the physician and do not complain about the early negative effects.

4. The usual starting dose was 50 mg t.i.d., a fairly low dose. Larger doses are usually arrived at slowly in clinical trials, so that the established drug effect could have overcome the onset effect.

5. It is usually difficult to train patients for valid self-reports. All our Ss had training in introspection (12), so that even very weak drug doses are detectable in our data. The Ss could not fail to be aware of the strong effects of 150 mg of benzquinamide.

6. Since animal tests (6) had shown this to be a "tranquilizer," the investigators may have had a mental set to ignore occasional contrary findings.

## 2. *Validity of These Data*

Since 1954, when we adopted the methods used in this experiment, we have had approximately 14,000 individual-subject contacts (opportunities for conversation with Ss while in the laboratory). Using other drugs with equal or stronger steady-state sedation (meprobamate, Dramamine, and others), we have never encountered the kind of reaction we obtained from each of the first five Ss given 150 mg of benzquinamide. On a few occasions, under high doses of Dramamine, we have had anger; and under high doses of Benzedrine, apprehension; but not such a complex of negative feelings. Placebo reactions do not explain this, since they did not occur during the three or more placebo sessions of these Ss nor in their more numerous sessions with other drugs.

The steady-state data, based on 13 Ss given 75 mg, are on relatively few Ss (fewer than we ordinarily use), but this is probably enough. In more than 20 separate experiments on Dramamine and nearly as many on Benzedrine, using group sizes from 12 Ss to 41 Ss, we have always obtained consistent results. With a drug effect comparable in magnitude to that of the 75-mg dose of benzquinamide, we have observed that on any one administration (Dramamine, Benzedrine) about 86 per cent of the Ss will show the typical effect. With 100 mg of Dramamine this may rise to 100 per cent. It should also be noted that in the course of the present experiment these same Ss showed the characteristic reactions to Dramamine and Benzedrine and

lack of reaction to placebo; hence we trust these steady-state data on benzquinamide, especially in view of the internal consistency of the particular mood states that were affected.

### 3. *Mechanism of the Onset Effect*

To our knowledge, the mechanism of action of benzquinamide is unknown. Our data, showing onset effects characterized by feelings of restlessness, confusion, unreality, anger, apprehension, unsociability, depression, and sedation (with other feelings occasionally thrown in for good measure), suggest that many neurological-emotional systems are being stimulated simultaneously. It is also suggested that the steady state of sedation is established when some one brain-stem system suppresses the others. In the absence of more information, further speculation is not justified.

### 4. *Do the Onset Effects Have Diagnostic Uses?*

We suggest to psychiatrists that they try to exploit the anomalous onset effects of benzquinamide, administering it to become effective during a psychiatric interview, to find whether the stress of an onrush of negative emotions may not be helpful in understanding the patient's personality and problems. We believe that this action of benzquinamide gave us valuable insights into our Ss: for example, some turned their hostility inward; others, against their environment.

### 5. *Suggestions for Clinical Research in Psychopharmacology*

The normal techniques of clinical drug testing often seem unduly slow in arriving at a valid consensus of opinion. We have three suggestions for expediting an understanding of the action of new drugs with psychological effects.

1. Skilled professional people (the investigators) should try a new drug first on themselves, taking gradually increasing doses separated by two or more days. (In our experience in training coinvestigators to introspect regarding drug effects, we have found great differences in this ability; so not everyone can take this suggestion.)

2. In early clinical trials, patients should be under continuous observation and informal, relaxed social interaction with the investigator for two to three hours after taking a drug, in order to observe the transition into the drug state. This is equally important whether a drug does or does not have onset effects that are different from its steady-state effects.

3. There are important advantages in trials made with normal subjects,

handled in such a way as to eliminate the placebo effects and to maximize introspective sensitivity (12). While therapeutic questions often cannot be answered by such data, they are likely to yield many other returns not easily obtained from patients. (There appears to be a widespread tendency in drug research to use normal Ss mainly for performance tests, in spite of some 80 years of research demonstration that these have little to teach us.)

### G. SUMMARY

In an experiment on 50 normal young men, benzquinamide in various doses produced sudden onset effects with feelings of restlessness, confusion, unreality, apprehension, anger, unsociability, depression, and sedation. In 57 double-blind administrations this was known to occur 19 times. Its duration was brief and gave way to a steady state of sedation. Hypotheses are offered to explain the failure of clinical investigators to report such effects. It is suggested that the onset effects may have personality-diagnostic uses and that experiments on normal subjects are an important adjunct to clinical trials.

### REFERENCES

1. BISHOP, M. P., GALLANT, D. M., & STEELE, C. A. A controlled evaluation of benzquinamide: Behavioral toxicity with high dosage levels in schizophrenics. *Curr. Ther. Res.*, 1963, **5**, 238-244.
2. FELDMAN, P. E. Clinical evaluation of benzquinamide (P-2647) in the control of tension states and "hyper-syndromes." *Psychosomatics*, 1962, **3**, 1-4.
3. HASHKES, H. R. Evaluation and pattern response of a benzquinamide in 50 patients. *Clin. Med.*, 1962, **69**, 2239-2244.
4. NUSSBAUM, K., MCCUSKER, K., & KURLAND, A. A. Benzquinamide in treatment of the psychotic patient. *Curr. Ther. Res.*, 1962, **4**, 513-523.
5. OVERALL, J. E., HOLLISTER, L. E., BENNETT, J. L., SHELTON, J., & CAFFEY, E. M. Benzquinamide in newly admitted schizophrenics: A search for patients best treated with a specific drug. *Curr. Ther. Res.*, 1963, **5**, 335-342.
6. SCHARBINE, A., WEISSMAN, A., FINGER, K. F., DELAHUNT, C. S., CONSTANTINE, J. W., & SCHNEIDER, J. G. Benzquinamide: A new antianxiety drug. *J.A.M.A.*, 1963, **184**, 276-279.
7. SETTEL, E. A clinical evaluation of benzquinamide with observations on the total activity potential. *Clin. Med.*, 1963, **70**, 957-962.
8. SHULKIN, M. W., DYKJ, R., BODI, T., LIM, M., SEIGLER, P. E., MAPP, Y., & NODINE, J. H. Preliminary human bioassay of benzquinamide in ambulatory psychoneurotics. *Curr. Ther. Res.*, 1962, **4**, 285-294.
9. SMITH, M. E. A pilot study of P-2647, a new benzquinamide derivative, as a psychotherapeutic agent in 44 cases of overt anxiety. *Amer. J. Psychiat.*, 1962, **118**, 937-938.
10. ———. The effects of benzquinamide. *Dis. Nerv. Syst.*, 1963, **24**, 116-122.
11. WEISSMAN, A. & FINGER, K. F. Effects of benzquinamide on avoidance behavior and brain amine levels. *Biochem. Pharmacol.*, 1962, **11**, 871-880.
12. WENDT, G. R., & CAMERON, J. S. Chemical studies of behavior: V. Procedures in drug experimentation with college students. *J. of Psychol.*, 1961, **51**, 173-211.

13. WENDT, G. R., CAMERON, J. S., & SPECHT, P. G. Chemical studies of behavior: VI. Placebo and Dramamine as methodological controls, and effects on moods, emotions, and motivations. *J. of Psychol.*, 1962, **53**, 257-279.

*Department of Psychology*

*University of Rochester*

*River Campus Station*

*Rochester, New York 14627*



## A FURTHER ANALYSIS OF THE ROLE OF CREATIVE THINKING IN HIGH-SCHOOL ACHIEVEMENT\*<sup>1</sup>

*College of Education, Kent State University*

---

KAORU YAMAMOTO

---

### A. INTRODUCTION

In recent years, the role of creative thinking in school achievement has drawn increasing attention from researchers in education and psychology. For example, it has been found that, among both high-school students (3) and elementary-school pupils (6), those scoring in the top 20 per cent on intelligence tests (but not in the top 20 per cent on a battery of tests of creative-thinking abilities) and those scoring in the top 20 per cent on creative-thinking tests (but not in the top 20 per cent on intelligence) give comparable performance on standard achievement tests, in spite of more than 20-point differences in mean intelligence quotients between these two groups. As a result, an hypothesis has been presented that imaginativeness, curiosity, and similar qualities of highly creative subjects enable them to compensate for what they lack in memory and in other factors measured by a test of intelligence.

However, some doubts have been expressed on the generalization of the results obtained in these studies, because of the lack of representativeness of subjects. Both studies used selected populations in laboratory schools. For this reason, at least seven replications of the Getzels-Jackson study have been reported since 1959: four in elementary schools (2, 8), one in high school (10), one in a graduate course (8), and one in summer counseling institute (4).

In all of these studies, "highly intelligent" and "highly creative" groups were identified in essentially the same fashion as in the Getzels-Jackson study (3). There were no statistically significant differences on achievement measures in five schools out of the seven. The only exceptions were found in two of the four elementary schools but, even in those schools, it was observed that the highly creative groups tended to "overachieve" for their intelligence while the highly intelligent groups tended to "underachieve."

---

\* Received in the Editorial Office on May 15, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The study was supported by the Bureau of Educational Research, University of Minnesota.

Torrance (8) presented three possible explanations for the observed results. First, he mentioned the vastly different orientations towards learning in various schools. Second, agreeing with Anderson (1), he argued that there may be a cutoff point above which a higher *IQ* makes little difference and above which creative thinking abilities become important. Finally, he mentioned the necessity for considering different kinds of achievement and for developing measures to assess them.

One of many missing links in this chain of research is the answer to a simple question, "Do highly creative subjects perform better than low-creative subjects on standard achievement tests when the effects of intelligence are held constant for the two groups?" Unless the answer is "yes," one cannot assuredly insist that creative thinking plays an important role in school achievement, complementing or supplementing intelligence.

The foregoing question has been asked by Torrance (7). After having found, for 75 subjects in the fourth through sixth grades, statistically significant correlation coefficients (.37 to .53) between a total score on a creativity battery and standard achievement measures, Torrance partialled out the effects of intelligence to see to what extent creativity contributed to school achievement. The relationships between creativity scores and achievement measures remained statistically significant, ranging from .23 to .48. Torrance also noted that partialling out the effects of intelligence tended to raise the correlations in the fifth and sixth grades, but did not improve the relationships in the fourth grade.

In the present study, the author tried to obtain additional information to answer this simple but important question about the role of creative thinking in school achievement. The specific hypothesis tested was as follows: When effects of intelligence are fixed, highly creative *Ss* perform better than low-creative *Ss* on standard achievement tests, regardless of the subject matter involved.

## B. PROCEDURES

### 1. Subjects

The *Ss* of this study were 272 students of the ninth through twelfth grades of the University of Minnesota High School during the academic year 1959-1960. These students comprised 89.8 per cent of the total enrollment (89.6 per cent of the boys and 89.9 per cent of the girls) of the school in these grades. Complete data were not available for the remaining 10.2 per cent. From these 272 subjects, the following two subgroups were identified in each grade: (a) High-Creativity Group—54 *Ss* ranking in the upper 20 per cent

on the measure of creative thinking, and (b) Low-Creativity Group—54 Ss ranking in the lower 20 per cent on the measure of creative thinking. Table 1 presents certain details and Table 2 gives the mean creativity scores and mean IQs for the two subgroups identified.

TABLE 1  
ENROLLMENT, POPULATION, AND TWO SUBGROUPS BY GRADE AND SEX

Grade	Total enrollment*			Sample population			High Creativity			Low Creativity		
	T	B	G	T	B	G	T	B	G	T	B	G
9	76	38	38	70	35	35	14	3	11	14	9	5
10	78	39	39	69	35	34	13	9	4	13	10	3
11	77	41	36	68	34	34	14	4	10	14	9	5
12	72	36	36	65	34	31	13	5	8	13	11	2
Total	303	154	149	272	138	134	54	21	33	54	39	15

\* T = total, B = boys, G = girls.

TABLE 2  
MEANS AND STANDARD DEVIATIONS OF IQ AND CREATIVITY FOR POPULATION AND TWO SUBGROUPS

Variable and group	Boys	Mean Girls	Total	SD	Total Range
<i>IQ</i>					
Population			118.32	15.00	
High Creativity	129.90	122.79	125.56	13.75	79 to 150
Low Creativity	115.79	112.00	114.74	16.03	80 to 150
<i>Creativity</i>					
Population			236.77	67.93	
High Creativity	352.76	333.85	341.20	47.47	276 to 498
Low Creativity	158.33	160.13	158.83	25.80	94 to 198

## 2. Collection of Data

For each subject the following data were obtained.

a. *Intelligence quotient.* Intelligence quotients were derived from the verbal battery of the Lorge-Thorndike Intelligence Tests administered under the supervision of the school psychometrist as a part of the school's annual testing program.

b. *Creativity score.* In September 1959, the Test of Imagination and the Ask-and-Guess Test (9) were administered to all the subjects, in a group situation. The results were scored on Fluency, Flexibility, Adequacy, Cleverness, Inventivelevel, and Constructiveness; and the unweighted sum of all the subscores was used as a single index of creative thinking.

*c. Achievement score.* Achievement data were obtained from the school's official records of the Iowa Tests of Educational Development, which tests were administered as a part of the school's regular test program. Percentiles were translated into standard scores (zero to 30) according to the beginning-of-the-year grade-percentile norms based on the results of the 1942 Fall Testing Program of The Iowa Tests of Educational Development (5). Nine subscores, one on each of the nine tests, plus one composite score covering Tests 1 through 8, were derived: (a) Test 1—Understanding of Basic Social Concepts, (b) Test 2—General Background in the Natural Sciences, (c) Test 3—Correctness and Effectiveness of Expression, (d) Test 4—Ability to Do Quantitative Thinking, (e) Test 5—Ability to Interpret Reading Materials in the Social Studies, (f) Test 6—Ability to Interpret Reading Materials in the Natural Sciences, (g) Test 7—Ability to Interpret Literary Materials, (h) Test 8—General Vocabulary, and (i) Test 9—Use of Sources of Information.

### C. RESULTS

The mean standard achievement scores for the two subgroups on The Iowa Tests of Educational Development are presented in Table 3.

TABLE 3  
MEAN STANDARD SCORES OF TWO SUBGROUPS ON THE IOWA TESTS OF  
EDUCATIONAL DEVELOPMENT

Subtest	High Creativity			Low Creativity		
	Boys	Girls	Total	Boys	Girls	Total
1	23.28	20.65	21.67	19.69	16.83	18.90
2	23.48	19.76	21.20	20.04	16.14	18.96
3	19.84	21.09	20.61	16.61	18.93	17.25
4	22.88	21.29	21.91	19.31	18.38	19.05
5	22.73	21.75	22.13	17.78	18.57	18.00
6	23.71	22.09	22.72	19.93	18.74	19.60
7	21.52	21.57	21.55	17.07	18.28	17.41
8	22.85	22.20	22.45	19.65	19.97	19.74
9	22.85	22.33	22.54	18.90	20.75	19.41
Composite	23.63	22.42	22.89	19.49	18.93	19.33

When a two-way analysis of covariance was applied to the scores, the results shown in Table 4 were obtained. The two dimensions in this analysis were subgroups and subtests.

From Table 4, it is apparent (a) that the difference between the two subgroups, High Creative and Low Creative, was statistically significant even when the effects of intelligence were fixed for these two groups; (b) that adjusting the means by the analysis-of-covariance technique decreased the

probability of the  $F$  ratio from .001 to .01; (c) that there was no difference between various subtests so far as  $Ss'$  achievement was concerned; and (d) that no interaction effect between grouping and subtest was detected.

In other words, the High-Creativity Group achieved significantly higher than did the Low-Creativity Group on all of the subtests of this achievement battery, even when the effects of intelligence were statistically fixed. Therefore, our hypothesis remains as it was.

TABLE 4  
ANALYSIS OF COVARIANCE OF STANDARD SCORES OF TWO SUBGROUPS ON THE IOWA  
TESTS OF EDUCATIONAL DEVELOPMENT

Source of variance	Original		Adjusted		Mean squares	$F$
	Degrees of freedom	Sum of squares	Degrees of freedom	Sum of squares		
Group	1	2731.75**	1	184.60	184.60	8.37*
Test	9	554.27	9	554.27	61.59	2.79
Group $\times$ Test	9	128.36	9	128.36	14.26	< 1
Residual	1060	34815.59	1059	23356.24	22.05	

\*  $p < .01$ .

\*\*  $p < .001$ .

#### D. DISCUSSION

The specificity of the student population studied still allows some doubts about the generalizability of the results. It should also be recognized that, by comparing two extreme creativity groups and thereby limiting the range of score variation, the covariance approach might have underplayed the contribution of intelligence to achievement, thus favoring the positive finding.

Within the confines of these and other limitations, the results suggest that performance on creative-thinking tests is related to success in school learning. Nevertheless, level of intelligence still seems to be a key factor; so a threshold of intelligence (1) beyond which high intelligence has relatively little significance for performance somehow should be determined. Torrance (8) has proposed an  $IQ$  of 120 as this cutoff point. In this connection it is interesting to note that, in the present study, achievement scores were adjusted to the average  $IQ$  level of the subgroups, which level was 120. At this level of intelligence, therefore, high creativity would seem to add to school achievement. A recent study (11) also attests to this.

The question of "how" (10) remains crucial in understanding the role that creative thinking plays in school achievement. This question is important not only for highly creative students with above-threshold intelligence, but also for those whose intelligence falls below this. Because the majority of the



school population is neither highly intelligent nor highly creative, and because education is to be equally concerned with all children (in contrast with a selected few), explorations into the process (or mechanism) of school achievement from the new perspective of creative thinking is urgent. In this connection, a careful re-examination of current standard achievement tests is also indicated.

#### E. SUMMARY

A total of 272 Ss of the ninth through twelfth grades of the University of Minnesota High School were administered the Minnesota Tests of Creative Thinking, the verbal battery of the Lorge-Thorndike Intelligence Tests and The Iowa Tests of Educational Development. Two subgroups were identified in each grade: (a) High-Creativity Group—a group in the upper 20 per cent on a test of creative thinking, and (b) Low-Creativity Group—a group in the lower 20 per cent on this same test. Fifty-four subjects were placed in each of these groups.

When the effects of intelligence were fixed through the analysis-of-covariance techniques, it was found that the High-Creativity Group performed significantly better than did the Low-Creativity Group on all of the subtests of this achievement battery.

As the mean *IQ* to which achievement scores were adjusted was about 120, the results lend some indirect support to Torrance's suggestion that a cutoff point might be somewhere around an *IQ* of 120.

#### REFERENCES

1. ANDERSON, J. E. The nature of abilities. In Torrance, E. P. (Ed.), *Education and Talent*. Minneapolis: Univ. Minnesota Press, 1960. Pp. 9-31.
2. BROWN, P. A. A study of highly creative and highly intelligent children on school achievement. Unpublished Master's paper, Marquette University, Milwaukee, 1960.
3. GETZELS, J. W., & JACKSON, P. W. *Creativity and Intelligence*. New York: Wiley, 1962.
4. PALM, H. J. An Analysis of Test-Score Differences Between Highly Creative and High Miller Analogies Members of the Summer Guidance Institute. Minneapolis: Bur. Educ. Res., Univ. Minnesota, 1959.
5. SCIENCE RESEARCH ASSOCIATES. *The Iowa Tests of Educational Development*. Chicago: Sci. Res. Associates, 1949.
6. TORRANCE, E. P. Explorations in Creative Thinking in the Early School Years: VI. Highly Intelligent and Highly Creative Children in a Laboratory School. Minneapolis: Bur. Educ. Res., Univ. Minnesota, 1959.
7. ———. Explorations in Creative Thinking in the Early School Years: VIII. I.Q. and Creativity in School Achievement. Minneapolis: Bur. Educ. Res., Univ. Minnesota, 1959.
8. ———. Educational Achievement of the Highly Intelligent and the Highly

Creative: Eight Partial Replications of the Getzels-Jackson Study. Minneapolis: Bur. Educ. Res., Univ. Minnesota, 1960.

9. ———. Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice Hall, 1962.
10. YAMAMOTO, K. The role of creative thinking and intelligence in high school achievement. *Psychol. Rep.*, 1964, **14**, 783-789.
11. ———. Threshold of intelligence in academic achievement of highly creative students. *J. Exper. Educ.*, 1964, **32**, 401-405.

*College of Education*  
*Kent State University*  
*Kent, Ohio 44240*

## EVALUATION OF SOME CREATIVITY MEASURES IN A HIGH SCHOOL WITH PEER NOMINATIONS AS CRITERIA\*<sup>1</sup>

*College of Education, Kent State University*

---

KAORU YAMAMOTO

---

### A. INTRODUCTION

As reviewed elsewhere, recent investigations in the area of creative thinking have been beset by ambiguous definitions, taken-for-granted assumptions, and the absence of valid and reliable instruments of measurement (4). Possibly, in their enthusiastic haste to explore the new frontier of our knowledge, investigators have spent comparatively little time in scrutinizing the criteria of creativity and in checking their instruments against available, though imperfect, criteria.

The study reported in this paper summarizes certain information on the "validity" of some of the creativity measures, as determined by peer nominations in a high school. A different aspect of this study (the relations between friendship nominations and creativity measures) has been reported in an earlier paper (5). The present paper is concerned with the relationships between some of the creativity measures and nominations on the corresponding, specific criteria of creative thinking.

### B. PROCEDURES

#### 1. *Subjects*

The Ss of this study were University of Minnesota High School students during the academic year 1959-1960. They were 428 in number (219 males and 209 females) and constituted 93.2 per cent of the total enrollment of the school (93.6 per cent of the males and 92.9 per cent of the females). Complete data were not available for the remaining 6.8 per cent of the students. Table 1 presents the number of subjects by grade and by sex.

#### 2. *Collection of Data*

The following data were obtained for each subject:

a. *Creativity score.* In September, 1959, the Test of Imagination and the

---

\* Received in the Editorial Office on May 18, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The study was supported by the Bureau of Educational Research, University of Minnesota.

TABLE 1  
Ss OF THE STUDY\*

Grade	Enrollment			Ss		
	T	M	F	T	M	F
7	80	40	40	58	29	29
8	76	40	36	76	40	36
9	76	38	38	76	38	38
10	78	39	39	78	39	39
11	77	41	36	71	37	34
12	72	36	36	69	36	33
Total	459	234	225	428	219	209

\* T = total, M = males, F = females.

Ask-and-Guess Test (3) were administered to all the subjects in a group situation. Protocols were scored for Fluency (both tests), Flexibility (Test of Imagination only), and Inventivelevel (Test of Imagination only).

b. *Peer nominations.* Early in June, 1960, a six-item sociometric instrument, "Who Does It?" was administered to all subjects. In answer to each of six questions, this form asked each subject to list three students from among those at his own grade level. The questions were as follows:

1. Who in this group (your class) comes up with the *most ideas*?
2. Who has the *most original or unusual ideas*?
3. If the situation changed or if a solution to a problem wouldn't work, who in your group would be the *first to find a new way of meeting the problem*?
4. Who in your group do the *most inventing and developing of new ideas, gadgets, and the like*?
5. Who in your group are *best at thinking of all of the details involved in working out a new idea and thinking of all of the consequences*?
6. Who are your *best friends* in the group?

For the purposes of the present study, nominations given to the first, third, and fourth questions were used as criteria for scores on Fluency, Flexibility, and Inventivelevel. For each subject, the number of nominations received on the respective dimensions was counted and recorded as his "nomination score." No weighting was made of the three names given on the nomination form.

## C. RESULTS

### 1. *Fluency*

Fluency scores on the Ask-and-Guess Test and on the Test of Imagination were independently correlated with Fluency-nomination scores. The results are presented in Table 2.

TABLE 2  
CORRELATION COEFFICIENTS BETWEEN FLUENCY SCORES ON THE ASK-AND-GUESS TEST AND  
THE TEST OF IMAGINATION AND FLUENCY-NOMINATION SCORES FOR EACH  
GRADE AND SEX

Grade	Sex	Ask-and-Guess Test		Test of Imagination	
		<i>r</i>	<i>z</i> (sex difference)	<i>r</i>	<i>z</i> (sex difference)
7	Males	.19		.10	
	Females	-.04	.816	.48**	1.255
	Both	.12		.14	
8	Males	.40**		.25	
	Females	.31†	.299	.34*	.353
	Both	.39***		.31**	
9	Males	.34*		.23	
	Females	.23	.409	.30*	.250
	Both	.30**		.25*	
10	Males	.59***		.62***	
	Females	.21	1.368	.29†	1.163
	Both	.41***		.47***	
11	Males	.11		.06	
	Females	-.08	1.147	-.18	.500
	Both	.02		-.05	
12	Males	.38*		.30†	
	Females	.11	.984	.30	.004
	Both	.18		.19	
Total					
	Males	.32***		.28***	
	Females	.16*	1.543	.19**	.921
	Both	.24***		.24***	

\*\*\*  $p < .001$ .

\*\*  $p < .01$ .

\*  $p < .05$ .

†  $p < .10$ .

Table 2 shows that, for both tests, the correlations vary in size from grade to grade. Although a few of the correlations are statistically significant, the number does not exceed that expected by pure chance. In no case, is there a statistically significant difference in correlations between males and females.

The two sets of correlation coefficients are comparable and reflect a high degree of correlation between the two Fluency scores (ranging from .67 to .95 for the six grades and .81 for the total group). In both sets, the correlation between the pencil-and-paper scores and the nomination scores is lowest in the seventh and eleventh grades. The low correlation coefficients at the seventh grade might be attributed to the relatively short peer acquaintance among subjects, but those at the eleventh grade are difficult to interpret. This



might be a mere artifact of this particular group or of these particular instruments, but it might also be suggestive of some kind of transition at this grade level.

## 2. Flexibility

The Flexibility scores on the Test of Imagination were correlated with the Flexibility-nomination scores, and the results are presented in Table 3.

TABLE 3  
CORRELATION COEFFICIENTS BETWEEN FLEXIBILITY SCORE ON TEST OF IMAGINATION  
AND FLEXIBILITY-NOMINATION SCORES FOR EACH GRADE AND SEX

Grade	Sex	<i>r</i>	<i>z</i> (sex difference)
7	Male	.04	.061
	Female	.06	
	Both	— .01	
8	Male	.21	.228
	Female	.27	
	Both	.26*	
9	Male	— .02	1.457
	Female	.35*	
	Both	.25*	
10	Male	.65***	.533
	Female	.48**	
	Both	.48***	
11	Male	.08	.547
	Female	— .05	
	Both	.00	
12	Male	.27	.617
	Female	.45**	
	Both	.19	
Total			
	Male	.22**	.098
	Female	.23***	
	Both	.22**	

\*\*\*  $p < .001$ .

\*\*  $p < .01$ .

\*  $p < .05$ .

It is seen in Table 3 that the variation of coefficients from one grade to another is marked, with recurrent pattern of the highest correlation in the tenth grade followed by the lowest in the eleventh grade being revealed again. Sex difference does not reach significance in any grade. The number of significant correlations over the six grades does not exceed that expected purely by chance for both sexes.

## D. DISCUSSION

The results obtained for the three scores, Fluency, Flexibility, and Inventivelevel, paralleled each other closely. This fact came as no surprise when it was learned that these three measures, both on the pencil-and-paper tests and on the peer nominations, tended to be interrelated to a high degree. For one thing, the correlation between the Fluency scores and the Flexibility scores on the Test of Imagination ranged from .79 to .99 among the six grades (.91 for the total group); that between Flexibility and Inventivelevel, from .56 to .84 (.79 for the total); and that between Fluency and Inventivelevel, from .50 to .86 (.59 for the total). The high correlations between the Fluency scores from the Ask-and-Guess Test and the Test of Imagination have already been mentioned.

In this connection, a recent factor-analytic study by Bowers (1) seems relevant. In factor analyzing the test results obtained from 203 Ss in the fourth, fifth, and sixth grades, Bowers found a massive group factor (called Factor A), common to all measures used, including the Ask-and-Guess Test and the Test of Imagination, accounting for a very great proportion of the variance in these variables. Fluency, Flexibility, and Originality scores derived from the Test of Imagination had the largest loadings in this factor, while the loading of the Ask-and-Guess Test measures (Fluency and Adequacy), and of the Inventivelevel score from the Test of Imagination, were substantial but lower.

As for the peer-nomination measures in the present study, the correlation between Fluency and Flexibility ranged from .74 to .92 over the six grades (.82 for the total group); that between Flexibility and Inventivelevel, from .35 to .72 (.62 for the total); and that between Inventivelevel and Fluency, from .38 to .74 (.65 for the total). Thus it would seem that there were considerable halo effects affecting these peer nominations. Possibly the Ss were not reacting in a discriminative fashion to the three questions involved. Wordings of these questions should be examined carefully for future studies because it is entirely possible that "creativity" represents a unitary concept in the lay mind and, therefore, all three questions tend to be interpreted as repetitions of a single question.

In working with tests of creative thinking, it seems to be a common procedure (with due skepticism) to obtain a composite score by simply summing all subscores on various dimensions. This composite score is then utilized as a single index of creativity. This procedure invites a question as to whether such a composite score yields more valid results than other scores when corre-

lated with a criterion. In the present study, correlations were computed between a composite test score (the sum of Fluency and Inventivelevel on the Test of Imagination) and a composite nomination score (Flexibility and Inventivelevel). Both of these composites consisted of the lowest-correlated scores among the respective component scores.

The resultant correlation coefficients ranged from .00 to .43 for male (.24 for the total) and from -.14 to .42 for female (.21 for the total). For the total groups, the correlation ranged from .00 to .41 over the six grades (.21 for the grand total). Sex differences were again nonsignificant. Generally speaking, then, combining subscores into a single composite score does not add much to the predictive efficiency of creativity measures, though in some grades a little gain can be noticed.

Some improved scoring schemes that are not totally dependent upon the "frequency" concept of creativity or originality are thus strongly indicated. The possibility of finding suppressor measures might also be considered. A single scoring scheme applied indiscriminately to all grades and sexes, without detailed consideration of the subtle psychological and social meanings of each product to a particular *S*, might in part be introducing some unspecified "penalizing or favoring" effect to the obtained results. Cultural expectations and social conditioning in terms of the boy's and the girl's "proper" behavior patterns are among the most important research areas in connection with the studies on creative thinking. Intensive studies of a small number of selected subjects, together with large-scale normative studies, are urgently needed to gain more insight into the process of man's imaginative behaviors.

In general, the values of correlation coefficients between the pencil-and-paper and nomination measures were around .20 and this corresponds well with the results of other validation studies (*cf.* 2). It was shown again that subjective judgments of creativity by untrained raters tend to be hampered by halo effects and other unexamined cultural expectations. Unless dimensions are sufficiently clarified to assure independent and fair judgment and unless judges know precisely what they are looking for, this kind of evaluation would seem to be unsatisfactory as an intermediate criterion of creative thinking. In addition, sociometric peer nominations as used in this study have a weak point in that the scores obtained are not only restricted in range, but also definitely skewed to the upper end, and fine discriminations are impossible—especially among subjects with a nomination score of zero. Possibly, some straightforward rating or ranking is preferable and some study with a decreased number of intimate peers is suggested.

The redundancy among the derived scores has been pointed out. It seems

necessary, before extending our research efforts into various luring paths, that we take stock of our present instruments and develop additional *complementary* measures of creative thinking. This remains as a laborious but fundamental task.

### E. SUMMARY

A total of 428 subjects of the seventh through twelfth grades of the University of Minnesota High School were administered (a) two tests of creative thinking—the Test of Imagination and the Ask-and-Guess Test—and (b) a peer-nomination form—“Who Does It?” From both of these, three scores—Fluency, Flexibility, and Inventivelevel—were derived.

When the scores on creativity tests were correlated with corresponding peer-nomination scores as the criteria, validity coefficients ranging from  $-.18$  to  $.65$  were obtained. Marked variations from measure to measure and from grade to grade were observed. For the total male sample, coefficients ranged from  $.20$  to  $.32$ , while they ranged from  $.16$  to  $.23$  for the total female sample. For the entire group, the values varied from  $.15$  to  $.24$ . Sex differences were statistically nonsignificant.

Three scores, Fluency, Flexibility, and Inventivelevel, were correlated with each other for both the Test of Imagination and peer nominations and it was found that these were highly intercorrelated among themselves. For the Test of Imagination, the highest correlation for the total was that between Fluency and Flexibility ( $.91$ ), the lowest, between Inventivelevel and Fluency ( $.59$ ). For the peer-nomination scores, the highest was that between Fluency and Flexibility ( $.82$ ); and the lowest, between Inventivelevel and Flexibility ( $.62$ ). Next, the Fluency and Inventivelevel scores on the Test of Imagination were summed to form a composite test score, and a criterion composite score was derived by combining the Flexibility and Inventivelevel-nomination scores. When these two composite scores were correlated, the coefficients ranged from  $-.14$  to  $.43$ . Variation among grades was again large.

Some discussion was presented on shortcomings of both the predictor and the criterion measures, and suggestions were made for further studies.

### REFERENCES

1. BOWERS, J. E. Explorations of Creative Thinking in the Early School Years: XIV. A Preliminary Factor-Analytic Study of the Creative Thinking Abilities of Children. Minneapolis: Bur. Educ. Res. Univ. Minnesota, 1960.
2. TAYLOR, C. W., & BARRON, F. Scientific Creativity. New York: Wiley, 1963.
3. TORRANCE, E. P. Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

4. YAMAMOTO, K. Creative thinking: Some thoughts on research. *Except. Child.*, 1964, 30, 403-410.
5. ———. Creativity and sociometric choice among adolescents. *J. Soc. Psychol.*, in press.

*College of Education*  
*Kent State University*  
*Kent, Ohio 44240*



## CODE LEARNING AND COLOR\*

*Department of Psychology, Miami University*

---

C. W. CRANNELL

---

### A. INTRODUCTION

It was recently brought to the attention of the author that some data, which he has casually collected, may have significance for certain practical problems relevant to information processing.<sup>1</sup> Specifically, situations are common in which many items of information, gathered and presented electronically, are fed to a human observer, who must make a decision as to the control of movements of some vehicle or perhaps decide (on the basis of the information received) whether or not to act. If the information being received is in the form of visual displays and if such displays are relatively complex, it would seem almost self-evident that use of color as a means of facilitating discrimination among the elements would be not only desirable but also necessary. But the data to be reported cast some doubt on carrying too far a generalization about the value of color in informational displays.

For a number of years, the author experimented with a variety of materials for use in comparison of the part and the whole methods for the learning of code. The model of the experiment originated with use of Morse code as material, but after World War II too many students were found to have basic familiarity with this code. Among the materials tried as substitutes and found wanting were geometric forms, playing cards, and Roman numerals—in each case randomly assigned to letters of the alphabet. It was found that the average student could not be expected to master such a code of 26 items within a 90-minute period. Then the writer happened upon a collection of Dennison decorative stickers representing flowers, birds, wild and domestic animals, seasonal features, and the like. A preliminary study was conducted using 20 such stickers assigned at random to 20 letters of the alphabet. It was found that, after a single presentation, some of the more rapid learners could correctly give the letter attached to each symbol. It became clear that

---

\* Received in the Editorial Office on May 18, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The writer is grateful to Anthony Debons, University of Dayton, for making clear the usefulness of the data, and to the late Warner Brown, University of California, in whose laboratory the methodology of the experiments was learned.

these colorful stickers offered material for very rapid learning of arbitrary associations.

## B. MATERIALS AND PROCEDURE

A collection of 40 easily distinguishable Dennison stickers was assembled. To each one was assigned a letter of the alphabet, a number from one to nine, or one of the words "zero," "yes," "no," "and," and "or," with the restriction that no pair would involve a *direct* association. As an example, a picture of a rose could not be assigned the letter "r." The stickers were pasted on 3" × 5" cards, with the appropriate letter, numeral, or word on the reverse side. Several sets of these 40 items were made. A set was also carefully photographed and several copies were made in black-and-white.

The experiment was conducted as a laboratory exercise in an upper-division course in experimental psychology. When the students assembled, they were randomly assigned positions as *E* and *S*, paired in teams, and assigned either the part or the whole method. All of them were then instructed as to procedure. In the whole method, *E* twice presented the entire set of 40 items, holding up each card in turn for three seconds and saying its appropriate designation: "a," "b," "c," and so on to the last item. After the second presentation, *E* presented the symbols from a shuffled set, holding up each card until *S* gave the correct response or an erroneous response or until five seconds had elapsed. In each case, *E* informed *S* of the correct response. When all 40 items had been tested, *E* again presented them in order and tested once more with a new shuffled order. This alternation of presentation and testing was continued until *S* had achieved three consecutive errorless trials.

The part method differed in that, after one presentation all the way through, *E* again presented only the first 10 items (a through j) and then tested *S* for the recall of these 10. This was done three times regardless of the correctness of *S*'s responses. Then *E* proceeded to present the next 10 items (k through t) and to test these 10 in the same manner. After part trials (in this way) for each of the remaining sets of 10 items, *E* presented and tested the whole series of 40 items (as in the whole method) until *S* reached the criterion of three consecutive errorless trials.<sup>2</sup>

<sup>2</sup> If an individual sets out to learn a poem or other material, he customarily surveys it by reading all the way through. He then proceeds to learn, using the whole or the part method as his inclination or training may dictate. If he settles upon the part method, he is ultimately confronted with the task of reciting the entire lesson. The purpose of the first "survey" presentation in both methods was to keep the experiment in line with practical learning procedures, with the use of precisely three "readings and recitals" in each part section being a concession to exactitude in experimentation.

Next, the teams were assigned to tables separated by screens, were given either the color code or the black-and-white code materials, and were kept under observation by the present writer or by his assistants. Any cases of irregular or inadequate performance by *Es* were eliminated from the data.

During a class period 19 to 21 days after the learning experiment, an opaque projector was set up, answer sheets were provided and, without any previous announcement, a recall test was performed. In the recall test, both the original stickers and their photographed reproductions were presented to all *Ss*. The order of presentation was random, with the restriction that the same symbol in its original color and in black-and-white could not be immediately consecutive.

### C. RESULTS

Table 1 shows the number of cases, means, and standard deviations obtained in original learning and in later recall for each of the four groups. The number of cases for recall data tends to be slightly smaller than the original number. In part, this reduction was due to absenteeism; in part, to the elimination of any *S* who may have had access to the experimental materials during the three weeks. The large difference between errors for the part and the whole methods is, of course, a function of the procedure used; but there is no significant effect either on trials to learn or on recall.

Within methods there is no significant difference between any pair of means. No increase in significance may be obtained by combining data for methods because the obtained differences are as often in favor of one code type as of the other.

In passing, it may be pointed out that recall after three weeks tends to exceed 50 per cent of the material. The range for individual *Ss* is from three to 40 items. For associations of a wholly arbitrary type, this seems to be effective retention. In fact, preliminary studies indicated that when there are only 28 items in the code, about 75 per cent retention is found after three weeks; while, for codes of such materials as geometric forms, Roman numerals, or playing cards, retention tends to be around 40 per cent.

### D. DISCUSSION

The negative results of this study suggest a cautious approach to the introduction of color in complex displays as a means of facilitating either learning or recall of the information to be received. Not covered by this study is the possibility that color may be of considerable value in cases in which learning or retention is relatively difficult. However, if the learning or re-

TABLE 1  
MEANS AND STANDARD DEVIATIONS IN ORIGINAL LEARNING AND RECALL AFTER THREE WEEKS

Method	Code type	N	Trials		Original learning		Errors	s	N	Number correctly recalled		Black-and-white
			M	s	M	s				Color	s	
Whole	Color	40	6.48	3.23	35.68	26.56		37	23.19	7.76	21.35	7.17
Whole	B-and-W	31	6.23	2.05	29.19	17.97		30	22.77	8.08	22.73	7.86
Part	Color	33	6.75	0.90	8.70	6.01		31	20.90	10.19	19.32	10.06
Part	B-and-W	28	7.00	1.16	9.00	8.33		26	23.46	8.96	24.15	8.89

tention is made easy, differentiation on the basis of color may not be important. The practical-information situation should aim, of course, in the direction of making these tasks easy.

The items to be learned in this study were easily distinguishable pictures. They could be transmitted by some device, such as television, and could be decoded rapidly into language by an observer. It would not be difficult to select observers who could retain the code perfectly after a relatively short learning period, probably less than 30 minutes, if the code were reduced to 26 pictures and their associated letters. So far as learning or retention may be involved, nothing would be achieved by the expensive addition of colored transmission.

#### E. SUMMARY

Upon exposure of associated pictures of the decorative sticker variety, four groups of college students learned to produce letters of the alphabet or numerals or simple words. Two methods of learning (whole and part) and two types of pictures (in full color and in black-and-white) were assigned at random to Ss so as to make up the four groups. Neither method nor code type was found to affect the number of trials taken in original learning nor the amount of recall after three weeks without intervening practice. The materials tended to be unusually easy to learn and to retain, and it is suggested that when this is true, little is to be gained by the use of color in the stimulus pattern.

*Department of Psychology*  
*Miami University*  
*Oxford, Ohio*



## MEAN ASSOCIATIVE LATENCIES OF 200 CVC TRIGRAMS<sup>1</sup>

*Department of Psychology, University of Hawaii*

---

RONALD C. JOHNSON

---

### A. INTRODUCTION

In a previous paper (2), the author presented data showing the existence of a strong negative relation between the mean associative latency ( $L$ ) and the Archer (1) associative value ( $a$ ) of consonant-vowel-consonant (CVC) trigrams. When corrected for attenuation, the relation between  $L$  and  $a$ , for the 80 trigrams used in this previous study was  $-.80$ . The same trigrams were used in learning experiments, and  $L$  and  $a$  measures were approximately equal in their ability to predict rate of learning. The writer wished to replicate the previous study and also to obtain a large pool of trigrams for which  $L$  values were obtained under strictly comparable conditions, so that, on the basis of their latencies, trigrams could be selected for use in learning experiments. The purpose of the present paper is to present replication data regarding the relation of  $L$  to  $a$  and also to present mean  $L$  values for 200 CVC trigrams.

### B. METHOD

#### 1. *Subjects and Materials*

Eighty undergraduate students served as  $S$ s in the latency task. Forty  $S$ s responded to 98 randomly selected CVC trigrams. Forty other  $S$ s responded to 102 randomly selected CVC trigrams. (Selection was random *except* that  $E$  eliminated from consideration the 80 trigrams for which latencies previously had been obtained.)

#### 2. *Procedure*

Each trigram was typed on a 3"  $\times$  5" card. The deck of trigrams, 98 or 102 in number, was shuffled prior to presentation to each  $S$ .  $S$  was told that he was to respond with a word to each trigram; that, if the trigram was a word, he was to say that word; that, if the trigram was not a word but

---

\* Received in the Editorial Office on May 13, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by research grant MH 08236-01 from the National Institute of Mental Health. The writer wishes to express his gratitude to Mr. Marvin Rohrer, who assisted in obtaining the data.

sounded like a word, he was to say that word; and that, if it were neither a word nor sounded like a word, he was to say the first word of which he was reminded when presented with the trigram. If *S* did not make an association within 10 seconds, the response was scored as 10 seconds. A Hunter Kloc-Kounter was activated on trigram presentation. *S* responded into a microphone that stopped the timer. Latencies were recorded to 0.001 second.

### C. RESULTS

The two lists of trigrams did not differ significantly from one another in *L* or *a*, and so are combined. The mean latencies and Archer association values for the 200 trigrams are presented in Table 1.

Reliability figures and correlation coefficients were computed for the two groups separately. The correlation between individual CVC trigram latencies of odd- vs. even-number *Ss* in the first group, who responded to 98 trigrams, is .89; for the second group of *Ss*, who responded to 102 trigrams it is .84. Archer reports a reliability of .88 for the *a* measure. The correlation between *L* and *a* for the first list of trigrams is  $-.86$ . Corrected for attenuation, the correlation is increased to  $-.97$ . The correlation between *L* and *a* for the second group of trigrams is  $-.76$ . Corrected for attenuation, the correlation is increased to  $-.88$ . Each correlation is significant beyond the .001 level of confidence. Both correlations are higher than the one previously cited (2), probably because the testing reported herein took place in an isolated, sound-proof room, so that distractions were at a minimum, while the previous latency data were obtained under something less than optimal conditions.

It is of interest that latencies are so low. The *mean* latencies are seldom over two seconds, the usual exposure time in experiments in which *a* data are obtained. There are a number of factors that probably produce these unexpectedly low latencies. First, many *Ss* in *a* experiments may make their first association fairly early in the exposure interval, and yet in the analysis of results are lumped together with those *Ss* who make an association at the very end of the two-second exposure, since all that is asked *S* is whether or not he made an association in the exposure interval. Secondly, *Ss* in the present study knew that they were required to make an association, while *Ss* in *a* studies are given more permissive instructions. Both reliability figures and the correlations between *L* and *a* indicate that *Ss* do not respond to this demand in an indiscriminate or random fashion. Lastly, *Ss* in the present study were exposed to a relatively small number of trigrams, under individual test conditions. This may have kept *Ss*' interest and attention level high. The fact that there are a number of three-letter words—e.g., ZED ( $a = 31$ );

TABLE 1  
OBTAINED LATENCIES (L) AND ASSOCIATION VALUES (a) OF 200 CVC TRIGRAMS

SYLL	L	a	SYLL	L	a	SYLL	L	a	SYLL	L	a
BAD	1.14	99	JOT	1.26	98	PIX	1.41	94	VAF	2.01	22
BAW	2.00	84	JOV	1.41	63	POK	1.18	84	VAS	1.32	76
BEC	1.60	80	JUR	1.45	62	PUK	1.74	83	VAW	2.82	16
BEL	1.16	99	JYB	1.56	44	PUZ	1.88	54	VAY	1.70	43
BEX	2.44	31	JYD	2.03	11	PYK	1.61	58	VEK	2.39	27
BEZ	1.99	44	JYK	2.13	6	QAL	2.46	48	VIJ	3.25	9
BOP	1.38	98	KEG	1.62	99	QEC	1.88	16	VOK	1.93	42
BYG	1.16	49	KIG	1.54	30	QEK	2.19	17	VUT	2.28	29
CAF	1.26	90	KIL	1.23	92	QEN	1.51	31	VYK	2.10	39
CAG	1.67	54	KOT	1.63	86	QER	2.05	36	VYL	2.08	32
CEP	2.43	51	KOV	2.06	56	QIJ	2.20	5	VYT	2.61	15
CON	1.30	100	KUL	2.06	52	QOH	2.64	8	VYW	2.21	14
COP	1.37	100	KUT	1.64	86	QOK	2.27	15	WEJ	1.80	42
CUN	1.77	73	KYH	2.78	7	QUC	1.65	15	WEL	1.33	94
CYF	2.50	32	KYT	1.22	64	QUR	1.90	31	WEM	1.95	38
CYK	2.31	46	KYZ	2.33	18	QUZ	1.25	36	WUD	1.89	55
CYX	2.35	31	LAS	1.25	88	QYV	2.25	5	WYG	1.51	45
DAV	1.80	86	LET	1.04	100	RAF	1.52	81	WYL	1.54	49
DAY	1.08	100	LEW	1.87	92	RAM	1.09	100	XAZ	2.49	2
DEJ	2.96	13	LIC	1.33	89	REX	1.32	98	XEC	2.43	9
DON	1.21	100	LIV	1.14	93	RIQ	1.79	38	XEG	2.76	5
FEW	1.07	99	LOH	1.67	50	RIY	1.87	32	XEJ	3.28	3
FEZ	1.73	71	LOT	0.99	100	ROF	1.97	67	XIB	3.48	6
FYP	1.77	15	LUP	1.81	53	ROH	1.75	50	XIP	1.97	26
FYQ	2.03	10	LUT	1.78	54	ROJ	1.57	36	XIW	2.83	3
GAD	1.36	97	MAV	1.79	58	ROM	1.10	82	XOJ	2.91	2
GAH	1.73	35	MEL	1.35	97	ROW	1.12	99	XUF	2.51	3
GAN	1.87	61	MIJ	2.04	37	RUB	1.02	99	XUK	2.82	5
GAY	0.95	99	MIZ	2.16	58	RUQ	1.78	24	XUM	2.18	13
GIZ	1.97	31	MOS	1.31	75	RUT	1.26	99	XUY	2.78	3
GOX	2.38	12	MOV	1.11	78	SAW	1.02	100	XYH	2.98	1
GYC	2.33	10	MOZ	1.73	59	SEC	1.31	88	YAD	2.32	25
GYT	1.98	37	MUB	1.65	24	SEF	1.72	41	YAL	1.73	62
HAG	1.19	99	MUL	1.57	74	SEG	1.58	68	YAP	1.51	95
HAY	1.03	100	MYC	1.63	48	SER	1.14	87	YAW	1.56	76
HEG	1.96	45	MYF	2.18	25	SET	1.03	99	YES	0.92	99
HET	1.60	65	MYQ	2.27	26	SIP	1.04	99	YEW	1.91	80
HEW	1.98	94	MYX	1.83	47	SOJ	1.99	19	YIL	1.97	21
HEX	1.45	90	MYZ	2.03	23	SOX	1.12	96	YIT	2.42	35
HIB	1.94	44	NAS	1.76	55	SUQ	1.93	35	YUT	2.32	24
HIQ	1.98	47	NAW	1.70	94	SUV	2.05	22	ZAV	2.62	24
HOM	1.14	91	NAZ	1.86	51	SYL	1.82	69	ZAW	1.93	30
HYR	1.52	57	NEP	1.93	46	TAG	1.29	100	ZEN	1.92	38
JEG	1.75	29	NES	1.67	92	TAR	1.14	100	ZER	2.47	34
JEH	2.53	25	NIH	1.95	26	TIK	1.54	84	ZIB	2.60	28
JIG	1.21	98	NIJ	2.81	7	TOF	1.91	47	ZIL	2.07	37
JIS	1.91	50	NIZ	2.20	26	TOG	1.67	97	ZIS	2.19	33
JIV	1.53	52	NUG	1.45	66	TOH	1.61	46	ZUD	3.03	29
JOF	2.23	19	PAT	0.99	99	TOZ	1.94	40	ZUP	1.81	34
JOL	1.23	63	PIM	1.73	57	TUQ	1.54	33	ZYL	2.49	29

ZEN (38); FID (52); TAW (59); FEY (68); SOT (70); CAY, FEN, and SIB (74); FOP and BEY (75); NUB and YAW (76); YON (78); YEW (80); WOT and NIB (81); JIB (82); BOT and FOB (83); MOT and GYP (86); VOW (87); NIT and WAN (88); YIP (89)—with  $a$  values under 90 suggests that  $Ss$  may be somewhat inattentive in experiments aimed at determining  $a$ .

One might reasonably ask the question of why we need still another measure of the characteristics of CVC trigrams when we already have  $a$  data, measures of rated meaningfulness ( $m'$ ) and rated associations ( $a'$ ) as a result of Noble's work, (3) and pronounceability and frequency data from that of Underwood and Schulz (4). The data from Johnson (2) do not support the notion that  $L$  has any advantage, in terms of predictive power, over  $a$ . However, it would appear that  $L$  has one special advantage over  $a$ ; it is a continuous measure, to .001 second, while measures of  $a$ , despite the fact that many  $Ss$  are used, are essentially dichotomous (Do you have an association or don't you?) measures. It seems reasonable to believe that measures of degree are potentially more sensitive than measures of kind. Also, as noted in the previous study of  $L$ , latency may have an advantage over  $a$ ,  $a'$  or  $m'$  in that it ( $L$ ) is a term used as a measure of response strength in the larger areas of operant and respondent learning. Verbal learning has been a "tight little island," separate from both. Anything that can be done to bring verbal learning back to these larger areas may be valuable, since it then can be determined whether the same laws hold true across areas. A use of interchangeable terms may allow us, for example, to determine whether number of available responses and the strength of any given response are negatively related in the area of verbal behavior, as is generally the case in operant or respondent behavior, or whether the strengths of different associative responses summate, so that number and strength of responses are positively related, as seems likely to be the case in the area of verbal learning.

#### D. SUMMARY

Associative latency ( $L$ ) time for 200 CVC trigrams was obtained from 80  $Ss$ , 40 of whom responded to 98 trigrams, and 40 of whom responded to 102 trigrams. The correlations between  $L$  and association value ( $a$ ), corrected for attenuation, were  $-.97$  and  $-.88$ , respectively, for the two groups of trigrams. The potential value of the  $L$  measure was discussed.

#### REFERENCES

1. ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monog.*, 1960, 74, Whole No. 497.

2. JOHNSON, R. C. Latency and association value as predictors of rate of verbal learning. *J. Verb. Learn. & Verb. Behav.*, 1964, 3, 77-78.
3. NOBLE, C. E. Measurements of association value ( $a$ ), rated associations ( $a'$ ) and scaled meaningfulness ( $m'$ ) for the 2100 CVC combinations of the English alphabet. *Psychol. Rep.*, 1961, Monog. Suppl. 3-V 8.
4. UNDERWOOD, B. J., & SCHULTZ, R. W. Meaningfulness and Verbal Learning. New York: Lippincott, 1960.

*Department of Psychology*  
*University of Hawaii*  
*Honolulu 14, Hawaii*





## THE SOCIOMETRIC COMPOSITION OF THE GROUP AS A DETERMINANT OF YIELDING TO A DISTORTED NORM\*

*University of Minnesota and University of California*

---

FRANK B. W. HARPER AND READ D. TUDDENHAM

---

### A. INTRODUCTION

Previous research on yielding to a distorted norm has suggested that, even among groups of strangers, emotional attitudes of group members towards one another may be of significance in determining yielding. See Tuddenham, Macbride and Zahn (10), Tuddenham (7), and Tuddenham and Macbride (9). Of interest therefore is the extent to which yielding is affected by acquaintance of the group members with one another. Furthermore, yielding may be affected not only by knowing the others in the group, but also by whether one's feelings towards them are predominantly positive or negative. The present study investigates the extent to which acquaintance with others affects yielding behavior and tests the hypothesis that yielding is higher in a group composed of people who are mutual friends than in a group of people who are mutually indifferent or who dislike one another. The hypothesis is derived from theoretical considerations concerning the conformist effects of group attraction of members and is suggested by earlier work: e.g., that of Wilson (11) and of Jackson and Saltzstein (5). A group of mutual friends is an attraction group in the sense that each wants to belong and will seek spontaneously to join with the others in work or play groups, where joining is relevant. A judgment that is represented as the group consensus is assumed to have greater significance for each member of an attractive group than for members of groups whose mutual relationships are characterized either by indifference or negative feelings.

### B. METHOD

#### 1. Subjects

The subjects were 95 student nurses in four training classes at two different hospitals. A first-year class contained 40 members, two second-year classes contained 39 and 63 members respectively, and one third-year class contained 32 members. Members of the classes had known each other for periods of

---

\* Received in the Editorial Office on May 25, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

time varying from six months to two and one-half years. All members lived in residence, usually on the same floor. Some were roommates. There was ample opportunity in social and work settings for each student to make contact with all others in her class and to establish preferences and avoidances for certain of them. However, as would be expected in groups of the sizes indicated, a number of the relationships were casual and involved no significant mutual attraction or repulsion. Subjects were tested in groups of five, constituted according to criteria based on sociometric rating within the class concerned.

### 2. *Sociometric Choices*

Each girl filled out a sociometric form on which she ranked all members of her class in order of closeness of friendship to her, starting with her best friend and ending with the girl she least preferred in the class. Usually, subjects filled out the two extreme ends of the list first since they were able without much trouble to discriminate people they really liked and people they did not care for. All subjects' lists were transferred to a master sociometric matrix in which the rank assigned by each girl to each other girl and the rank assigned to each girl by each of the other girls were obtainable by inspection within cells. From this matrix two types of groups were chosen. The first type contained five girls who gave each other high mutual ranks on close friendships, and the second type contained girls who gave each other low mutual ranks on friendship. There were nine groups of close friends and 10 less-preferred groups. It was assumed that the groups of close friends functioned as attraction groups and that the groups of less-preferred girls functioned as nonattraction groups. From checks with other sources in the schools of nursing, it was clear that strong mutual dislikes and antagonisms were prevalent among several members of each of the less-preferred groups, though not every member actively disliked every other one.

### 3. *Apparatus*

The apparatus was that used and described in detail by Tuddenham (6). By means of lights that appeared on a panel in front of each subject, the basic design allowed the experimenters to supply false information to each of the five participants in the experiment without their knowing about it. The subjects were led to believe that, with regard to a set of simple perceptual problems projected on the wall facing the group, the lights indicated the judgments of the others in the group. In fact, the lights represented a pre-arranged pattern of answers simulated by the experimenter and originated at a switch panel in another room.

#### 4. *Items*

The items were a set of 46 visual-judgment problems, prepared as 2" by 2" slides. Each item constituted a nine-part multiple-choice problem with the alternatives ordered quantitatively and numbered from one to nine. On most trials, the choices simulated as coming from "the others" corresponded to the modal choices of a control group. However, on 14 critical trials distributed through the series, the simulated judgments were deliberately distorted. On such trials, the pseudoconsensus lay two steps in the series beyond the choice that corresponded to the 95th percentile of the response distribution of a control group answering without knowledge of other people's judgments. Answers distorted to such a degree are grossly and obviously wrong.

#### 5. *Procedure*

The subjects were instructed initially to be as accurate as they could be in making their judgments, and these instructions were repeated a second time after the initial practice session. During the presentation of the experimental series, on about every third trial, each subject was instructed to answer "last," and, on such trials, the group norms were distorted according to the procedure described. Both accurate and distorted group judgments were simulated on all orders of answering, but only those trials in which subjects answered last were used for calculating yielding scores.

#### 6. *Scoring*

The yielding score for each subject was computed by adding the assigned ordinal numbers of the alternatives chosen on critical trials, adjusted where necessary for the direction of yielding. The mean yielding scores for the separate groups and for the combined attraction and nonattraction groups were also computed.

### C. RESULTS

The mean yielding scores for the attraction groups and the nonattraction groups are shown in Table 1. For comparison, the mean yielding score of a comparable sample of strangers, as reported by Tuddenham (8), is also shown. Each of these means was very significantly different from the scores of a control group of women college students. See Tuddenham (8). As in Tuddenham's study, there was a significant yielding effect as a function of being exposed to the false norms.

The basic experimental hypothesis was that yielding would be higher in the attraction groups than in the nonattraction groups. A comparison of the

TABLE 1  
MEANS AND VARIANCES OF YIELDING SCORES OF ATTRACTION AND NONATTRACTION  
GROUPS, AND OF GROUPS OF STRANGERS

Parameter	Group			
	Attraction ( <i>N</i> = 45)	Nonattraction ( <i>N</i> = 50)	Strangers* ( <i>N</i> = 33)	Controls* ( <i>N</i> = 31)
Mean	63.16	63.00	60.14	46.82
Variance	103.09	172.90	56.40	3.50

\* Data for strangers, and for controls answering without knowledge of others' choices, are based on women college students. See Tuddenham (8, p. 235).

means of both types of groups showed that the average score of the attraction groups was 63.16, while the average score for the nonattraction groups was 63.00. The difference between these means is obviously nonsignificant, and the experimental hypothesis must be rejected. An overall *F*-test comparison showed furthermore that there was not only no difference between the attraction groups and the nonattraction groups, but also no difference between them (singly or combined) and the strangers' group.

#### D. DISCUSSION

The rejection of the experimental hypothesis indicated that yielding was unaffected by the emotional relationships within the groups. Yielding was about the same regardless of whether the group was composed of friends or not. Examination of individual subject's yielding scores revealed that there was considerable variability from one subject to another in the same group. In groups of close friends, one subject might be a low yielder, while the girls sitting beside her might be high yielders. Some subjects in the nonattraction groups were low yielders, while others were high yielders. It appears, therefore, that the important determinants of yielding were intraindividual rather than interindividual. Alternatively, the group factors that might significantly affect yielding were not those used as independent variables in this study.

The finding that group attraction versus nonattraction was an irrelevant variable in affecting yielding suggests that the general premise that a person is affected more by people he likes than by people he dislikes has to be taken as applying only in particular contexts. Downing (3) reported failure to secure greater induction of the autokinetic effect in cohesive groups and suggested that the power of the group to influence its members was probably restricted to those situations in which cultural influences permitted such influence. Those studies that have found favorable conformity effects as a function of group attraction have tended to use socially relevant dependent variables, such as opinions about labor disputes [see Festinger, Gerard, Hymovitch,



Kelley and Raven (4)] or productivity variables [see Berkowitz (1)], whereas of two experiments reporting clearly negative results, both used perceptual judgment tasks, one (Downing) the autokinetic effect (3) and the other (Bovard), estimation of the number of dots on a screen (2).

### E. SUMMARY

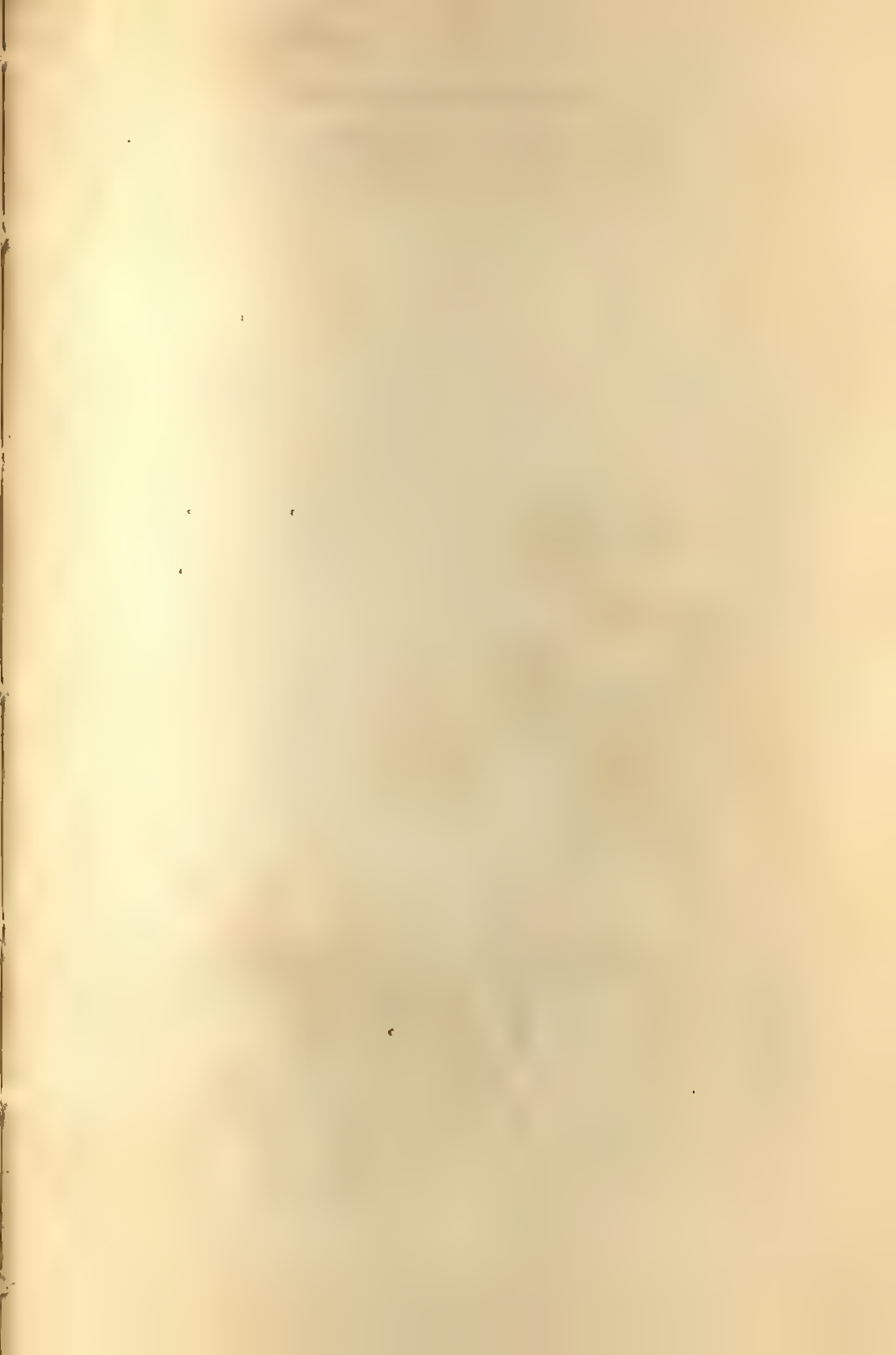
The hypothesis that a distorted group norm would secure greater yielding among members, as a function of whether it was interpreted as coming from a group of close friends or from a group of nonpreferred others, was tested by placing groups of student nurses in an Asch-type social-influence experiment. The results showed that yielding bore no relation to the type of group that subjects were in. Compared to a group of strangers, acquaintance groups were not significantly different in yielding behavior. Yielding appears to be intraindividually determined in large part, rather than being a function of friendship 'differences' or acquaintanceship differences. Yielding to a false group norm in experiments using perceptual problems may be unaffected by emotional relationships among group members, whereas socially relevant problems may be affected.

### REFERENCES

1. BERKOWITZ, L. Group standards, cohesiveness and productivity. *Hum. Relat.*, 1954, **7**, 509-519.
2. BOVARD, E. W. Conformity to social norms and attraction to the group. *Science*, 1953, **118**, 598-599.
3. DOWNING, J. Cohesiveness, perception and values. *Hum. Relat.*, 1958, **11**, 157-166.
4. FESTINGER, L., GERARD, H. B., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327-346.
5. JACKSON, J. M., & SALTZSTEIN, H. D. The effect of person group relations on conformity processes. *J. Abn. & Soc. Psychol.*, 1958, **57**, 17-24.
6. TUDDENHAM, R. D. The influence of a distorted group norm upon individual judgment. *J. of Psychol.*, 1958, **46**, 227-241.
7. ———. Correlates of yielding to a distorted group norm. *J. Personal*, 1959, **27**, 272-284.
8. ———. The influence of a distorted norm upon the judgments of adults and children. *J. of Psychol.*, 1961, **52**, 231-239.
9. TUDDENHAM, R. D., & MACBRIDE, P. The yielding experiment from the subject's point of view. *J. Personal*, 1959, **27**, 259-271.
10. TUDDENHAM, R. D., MACBRIDE, P., & ZAHN, V. The influence of the sex composition of the group on yielding to a distorted norm. *J. of Psychol.*, 1958, **46**, 243-251.
11. WILSON, R. S. Personality patterns, source attractiveness and conformity. *J. Personal*, 1960, **28**, 186-199.

*Institute of Child Development*  
*University of Minnesota*  
*Minneapolis 14, Minnesota*







## CONSISTENCY OF CROSS-TEST INDIVIDUAL DIFFERENCES IN PERSONALITY-INVENTORY FAKING\*

*Department of Psychology, University of Bridgeport*

---

JOHN R. BRAUN

---

### A. INTRODUCTION

In the course of investigations of the fakability of various forced-choice and disguised personality inventories, the author has noticed large individual differences in the ability of different Ss to change their scores when directed to try to do so (1, 2, 3, 4, 5, 6, 7). A question arises as to the generality of these individual differences. That is, will those persons who, under faking instructions, make large score changes on one inventory be able to perform in a comparable manner on a different inventory? Will there be some consistency of magnitude of score change across two different tests?

A related question was investigated by Garry (8), who dealt with the correlations between faking scores on different scales of the same test. He had college students take the Strong Vocational Interest Blank under standard directions and then under instructions to answer as would a physician, minister, lawyer, and president of a manufacturing concern. All of the intercorrelations among the faking scores were under .36, which fact Garry interpreted as showing that there is no marked general-faking ability. He felt, however, that the fact that nearly all of the correlations were positive suggested the possibility of some weak general-faking ability.

The present study investigated the consistency of individual differences in score-change magnitudes across two different forced-choice personality inventories.

### B. METHOD

Ss were students in undergraduate educational-psychology and social-psychology classes ( $N = 68$ ; 13 Male, 55 female). Inventories used were the Gordon Personal Inventory (GPI) and the Survey of Interpersonal Values (SIV). The GPI (9) has scales for cautiousness, original thinking, personal relations and vigor; while the SIV (10, 11) deals with support, conformity, recognition, independence, benevolence, and leadership. Both inventories have

---

\* Received in the Editorial Office on May 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.



been found by the author to be relatively susceptible to the effects of the top-management-executive faking set used in the present study (1, 3, 4).

During a regular hour-and-a-quarter class period *Ss* did the following: (a) took the GPI under instructions to try to appear the best-possible candidate for a top-management executive position (fake-good set), (b) took the SIV under a fake-good set, (c) retok the GPI under instructions to try to appear the poorest-possible candidate for the top-management executive position (fake-bad set), and (d) retok the SIV under a fake-bad set. For each *S*, the difference between the scores made on each scale under the opposed faking sets was computed, and these difference or change scores were totaled for each inventory. Thus for each *S* a GPI total faking score (representing the sum of the separate change scores for the cautiousness, original thinking, personal relations, and vigor scales) and an SIV total faking score (representing the sum of the separate change scores for the support, conformity, recognition, independence, benevolence, and leadership scales) were obtained. A Pearson product-moment correlation was then computed between these two inventory total change scores.

It should be noted that the foregoing procedure (involving administration of the two tests during the same class period) makes it possible for memory for specific responses to be a factor influencing the change scores. However, the particular order of test administration was such as to maximize retroactive-inhibition possibilities and thus makes it less likely that there would be a major influence from this source.

The rationale for the use of the difference score between fake-good and fake-bad administrations was as follows. There seemed to be no way to get meaningful absolute faking scores on the inventories employed. One might have added the scores for each *S* under the fake-good set on, let us say, the Survey of Interpersonal Values; but the interpretation of a composite score on support, conformity, recognition, independence, benevolence, and leadership seems questionable because some of these scales are negatively correlated.

Likewise, if the difference between regular-administration scores and fake-good scores were used, the *S*'s initial standing on each dimension would enter as a factor influencing the magnitude of the faking score. Thus if *Ss* tend to see executives as high on leadership, a person who scored low on this dimension artificially would get a higher faking score than one who scored high on it. Use of the difference between fake-good and fake-bad scores seems to avoid these difficulties.

## C. RESULTS AND DISCUSSION

The Pearson product-moment correlation between the GPI total faking scores and the SIV total faking scores was .35, which is significantly greater than zero at the .01 level. Thus greater-than-chance consistency of individual differences in score-change magnitudes across two different forced-choice personality inventories was found. Nothing can be inferred from this concerning possible consistency of faking *accuracy* across different inventories since the specific nature of the changes produced was not considered due to lack of data on the scores that top-management executives might make on these devices.

The two inventories used in this study are similar in format, and identical faking sets were used with each, thus conditions should be particularly favorable for demonstrating consistency of faking scores. The author hypothesizes that a design involving one test with a forced-choice format and another test with a true-false format, and a top-management-executive set used with one and a general "good-impression" set used with the other, would yield less evidence of consistency than the conditions of the present study. Therefore, although greater-than-chance consistency was found in the present study, the magnitude of the .35 correlation (accounting for only 12.28 per cent of the variance) is not impressive when considered in this light. The results of the present investigation are consistent with Garry's conclusion (8) that there seems to be no *marked* general-faking ability.

## D. SUMMARY

The consistency of faking ability across two forced-choice personality inventories was investigated using 68 undergraduates in psychology classes as Ss. They took the Gordon Personal Inventory and the Survey of Interpersonal Values under fake-good and then fake-bad top-management-executive faking instructions. The obtained Pearson product-moment correlation of .35 ( $p = .01$ ) between total faking scores on the two inventories is interpreted as being consistent with Garry's conclusion regarding absence of any marked general-faking ability.

## REFERENCES

1. BRAUN, J. R. Effects of a top management faking set on the Gordon Personal Inventory. *Psychol. Rep.*, 1962, **10**, 611-614.
2. ———. Differential susceptibility to faking of various Ghiselli Self-Description Inventory scales. *Psychol. Rep.*, 1962, **10**, 639-641.
3. ———. Effects of positive and negative faking sets on the Survey of Interpersonal Values. *Psychol. Rep.*, 1963, **13**, 171-173.

4. ———. Fakability of the Gordon Personal Inventory: Replication and extension. *J. of Psychol.*, 1963, **55**, 441-444.
5. ———. Effects of a "good-adjustment" faking set on the Holland Vocational Preference Inventory. *J. of Psychol.*, 1964, **57**, 303-306.
6. BRAUN, J. R., & ALEXANDER, S. Fakability of the Short Forced-Choice Anxiety Scale. *Psychol. Rep.*, 1961, **9**, 118.
7. BRAUN, J. R., & DUBÉ, C. S., II. Note on a faking study with the Famous Sayings Test. *Psychol. Rep.*, 1963, **13**, 878.
8. GARRY, R. Individual differences in ability to fake vocational interests. *J. Appl. Psychol.*, 1953, **37**, 33-37.
9. GORDON, L. V. Manual, Gordon Personal Inventory. Yonkers, N. Y.: World Book, 1956.
10. ———. Manual for Survey of Interpersonal Values. Chicago: Sci. Res. Associates, 1960.
11. GORDON, L. V., & MENSCH, I. N. Values of medical school students at different levels of training. *J. Educ. Psychol.*, 1962, **53**, 43-51.

*Department of Psychology*

*University of Bridgeport*

*Bridgeport 4, Connecticut*

## SHIFTS IN CONNOTATIVE MEANING OF WORDS AS A FUNCTION OF CONTEXT AND REINFORCEMENT\*<sup>1</sup>

New York University

M. E. TRESSELT AND M. S. MAYZNER

### A. INTRODUCTION

When an *S* is presented with a series of stimuli without any anchoring stimulus being given, and when that *S* is required to make categorical judgments of these stimuli, he will make his judgments along some subjective continuum. When the *S* is exposed to a limited portion of a stimulus range and then is given an expanded series of stimuli, a shift of judgmental categories has been demonstrated [e.g., Rogers (3), Tresselt (5)].

Most of the experiments made on the phenomenon of the shifting scale have utilized the extremes of the series of stimuli; for example, Mayzner and Tresselt (1) found that *Ss* who were asked to judge words, such as "table," "second," "turn," etc., for inclusion in the concept "economic," judged fewer such words as belonging to that concept if they were required first to make judgments for such words as "industry," "wages," "money," etc., than when there were no extreme or "strong" words given in the test situation.

Also studied has been the amount of learning on an original scale before the expanded series of stimuli are presented (1, 2, 5, 6, 7). In general, the results have indicated that the greater the experience with the prior range of stimuli, the greater the displacement of judgments in the new, expanded series of stimuli.

It is concern with these two variables that initiated the present investigation; that is, this study will examine systematically the anchoring effects of words that occupy different ranges of the continuum of words "belonging to the concept economic." For example, words at the high end of the scale will be mixed with words from the middle range of the scale; words at the low end of the scale will be mixed with words from the middle range; words intermediate between the high and the middle ranges will be mixed with words intermediate between the low and the middle ranges; words from the high, the middle, and the low ranges will be mixed; and finally words from

\* Received in the Editorial Office on May 28, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was supported in part by a grant made to M. E. Tresselt by the New York University Arts and Science Research Fund.

all ranges (i.e., high, intermediate, middle, intermediate, and low) will be mixed. Thus, the effects of anchoring on shifting absolute scale values will be studied over the entire range of the continuum.

In addition to studying anchoring effects by manipulating various word mixes or contexts, this study will attempt to determine if *S* can learn a category of "belongs" when there is no repetition of any single given word and also can learn a category "does not belong" even though the categories "does not belong" and "belongs" may be constituted of different ranges of the entire continuum. Such reinforcement, it is predicted, will magnify further the absolute-scale-value shifts obtained by manipulating only the context variables.

### B. SUBJECTS

The *Ss* were 650 students from introductory psychology classes at New York University, 50 in each of 13 different experimental conditions.

### C. STIMULUS MATERIALS

The stimulus material presented to each *S* consisted of a series of words, with each word being separately typed in capital letters on white 3"  $\times$  5" cards. The words were obtained from an earlier study (8) in which 1,000 words were presented to each of 200 *Ss*. One hundred *Ss* received the words in the order 1 through 1,000; the second 100 *Ss* received the words in the reverse order: 1,000 through 1. Each *S* was instructed to place a check mark next to each word that he judged as belonging to the concept "economic" and to leave blank each word that did not belong to the concept. For each of the 1,000 words judged, two "scale values" are available: i.e., (a) the number of *Ss* who judged a word as belonging to the concept, with the order 1 through 1,000; and (b) the number of *Ss* who judged a word as belonging to the concept, with the order 1,000 through 1. For 931 of the 1,000 words, the difference between the two frequency values was 10 or less; thus there is a reasonable degree of consistency between the two sets of scale values.

For this study, it was necessary to obtain word samples along the entire range of the scale: i.e., from words that few if any *Ss* judged as belonging to the concept "economic" to words that most or all *Ss* judged as belonging to it. Specifically, five category ranges were chosen (zero-to-20 per cent, 20-to-40 per cent, 40-to-60 per cent, 60-to-80 per cent, and 80-to-100 per cent). Words were then selected within each of these five ranges. For a word to be chosen, both of its scale values had to fall *within* the end points of each range. In the application of this criterion, it was found that only 66 words from the 1,000-word list fell within the 80-to-100 per cent range; therefore the same



number of words was selected for each of the other four ranges. For each range Table 1 presents 10 sample words and the two scale values associated with each word.

#### D. PROCEDURE

In order to examine the differential effects of context and reinforcement variables, 13 different experimental conditions were employed. The several conditions involved different mixes of the selected ranges and also conditions involving reinforcement. The experimental groups were as follows:

1. Ranges 40-to-60 per cent and 80-to-100 per cent intermixed.
2. Ranges 40-to-60 per cent and 80-to-100 per cent intermixed. *S* was told "right" if he judged an 80-to-100 per cent word as belonging to the concept or judged a 40-to-60 per cent word as not belonging. *S* was told "wrong" if he judged an 80-to-100 per cent word as not belonging to the concept or if he judged a 40-to-60 per cent word as belonging.
3. Ranges 20-to-40 per cent and 60-to-80 per cent intermixed.
4. Ranges 20-to-40 per cent and 60-to-80 per cent intermixed, with reinforcement as in Condition 2.
5. Ranges zero-to-20 per cent and 40-to-60 per cent intermixed.
6. Ranges zero-to-20 per cent and 40-to-60 per cent intermixed, with reinforcement as in Condition 2.
7. Range zero-to-20 per cent presented alone.
8. Range 20-to-40 per cent presented alone.
9. Range 40-to-60 per cent presented alone.
10. Range 60-to-80 per cent presented alone.
11. Range 80-to-100 per cent presented alone.
12. Ranges zero-to-20 per cent, 40-to-60 per cent, and 80-to-100 per cent intermixed.
13. Ranges zero-to-20 per cent, 20-to-40 per cent, 40-to-60 per cent, 60-to-80 per cent, and 80-to-100 per cent intermixed.

Each of the 650 *Ss* was assigned randomly to one of the 13 experimental conditions and was tested individually. In the conditions involving no reinforcement, the *S* was told that he would be presented with a series of words one at a time and that, for each word, he was to make a judgment as to whether or not the word belonged to the concept "economic." If he judged the word to belong, he was to say "yes"; if he judged the word not to belong, he was to say "no." *E* recorded the responses. For the reinforcement conditions (i.e., Conditions 2, 4, and 6) *S* was instructed as before; but in

TABLE 1  
SAMPLE WORDS AND ASSOCIATED SCALE VALUES (SV) FOR EACH OF THE FIVE RANGES EMPLOYED IN THE STUDY\*

0-to-20 per cent	SV	20-to-40 per cent	SV	40-to-60 per cent	SV	60-to-80 per cent	SV	80-to-100 per cent	SV
horizon	8-11	building	36-35	withdrawal	43-46	balance	78-74	investment	96-90
friend	2- 7	party	29-27	remit	42-42	borrow	72-74	tariff	85-86
object	13-18	coal	39-38	plan	44-49	coins	76-77	money	95-93
kiss	3- 1	city	23-25	personnel	53-46	deal	63-70	budget	93-88
boat	7-11	force	28-32	grain	48-44	margin	79-73	cost	90-90
day	6- 5	limit	38-36	foreman	47-43	fees	76-77	capital	90-93
blend	10-10	time	24-22	crops	54-47	gold	62-61	credit	94-87
library	7- 9	shop	33-29	luxury	55-57	trust	73-78	dollars	89-88
crime	19-17	load	30-28	railroad	48-52	partner	69-64	inflation	94-89
art	4- 4	local	30-29	upkeep	42-44	benefit	64-63	surplus	90-90

\* The first number following each word refers to the scale value (SV) obtained from the earlier scaling study (8) with the order 1 through 1,000. The second number refers to the scale value obtained with the order 1,000 through 1.

addition he was told that *E* would say "right" or "wrong" after each judgment and that *S* was to try to get as many judgments correct as possible.

Within each experimental condition, each *S* received the words in a "different" random order; however, in those conditions in which two ranges were mixed, successive blocks of 44 words always contained 22 words from one range randomly mixed with 22 words from the other range. In Experimental Conditions 12 and 13, successive blocks of 66 and 110 words, respectively, always contained 22 words from each of the three or five ranges. Further, the same 50 random orders generated for Experimental Conditions 1, 3, and 5 were used for Experimental Conditions 2, 4, and 6, so that the only difference between Condition 1 and Condition 2, between Condition 3 and Condition 4, and between Condition 5 and Condition 6 was the presence or absence of reinforcement for the *S*'s judgments.

#### • E. RESULTS AND DISCUSSION

To evaluate the results, two criterion measures were calculated. The first measure focuses on *S*'s performance; that is, for each range the number of words that *S* judged to belong to the concept "economic" was obtained and the mean of those values for the 50 *S*s was then calculated. The second measure focuses on word performance; that is, for each word the per cent of *S*s who judged the word as belonging to the concept was obtained and then the median of the 66 word scores was calculated. The results are presented in Table 2.

Because the major interests in this study are changes or shifts in the scale values of words as a function of context and reinforcement, the statistical analyses for the most part were applied to the word-performance scores rather than to the *S*-performance scores. However, Table 2 reveals that the directions of the changes or shifts in word scale values, as given by the median scores, are essentially in correspondence with the changes in *S*'s performance, as given by the mean scores. Since both mean and median scores reflect the same trends, all further discussion will be based on the median scores only, except for portions of Experimental Conditions 2 and 6 in which the mean scores are employed to assess the significance of changes or shifts in performance as they relate to reinforcement contingencies.

The Kolmogorov-Smirnov two-sample test (4) was employed to assess differences in median scale values between the experimental conditions and the various ranges. With the use of the two-sample test, one may examine deviations between the two cumulative distributions of word scale values for any particular pair of ranges that he wishes to compare. Almost without

TABLE 2  
MEAN NUMBER OF WORDS JUDGED TO BELONG TO THE CONCEPT "ECONOMIC" AND MEDIAN SCALE VALUES FOR THE WORDS IN EACH OF THE 13 EXPERIMENTAL CONDITIONS

Experimental condition	0-to-20			20-to-40			Ranges						60-to-80			80-to-100		
	per cent			per cent			40-to-60			per cent			per cent			per cent		
	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>M</i>	<i>SD</i>	<i>R</i>	<i>M</i>	<i>SD</i>	<i>R</i>	<i>M</i>	<i>SD</i>	<i>R</i>	<i>M</i>	<i>SD</i>	<i>R</i>	<i>M</i>	<i>SD</i>	<i>R</i>
1							40.36		62%							64.00		98%
2							16.73		36-92%							3.56		82-100%
3																		
4							16.22 <sup>a</sup>									59.14		91%
5							6.98 <sup>a</sup>									6.93		56-100%
6																56.08		88%
7, 8, 9																4.47		34-100%
10, 11																		
12																		
13																		

<sup>a</sup> No median scale values are given as a trend analysis revealed that performance was changing over successive 44-word blocks (7.26-4.78-4.18).

<sup>b</sup> No median scale values are given as a trend analysis revealed that performance was changing over successive 44-word blocks (5.36-4.18-3.72).

exception, all of the relevant comparisons yielded  $p$  values of less than .05.

The results presented in Table 2 reveal a strong and systematic effect both of context and of reinforcement on shifts in word scale values over the entire continuum of the scale. Because a great many comparisons are possible, those relating to context only will be presented first; they will be followed by considerations relating to reinforcement.

### 1. Context

Experimental Conditions 12 and 13 constitute reference conditions against which the other conditions may be examined. First, however, Experimental Condition 13 needs to be related to the original scaling study (8) from which the words employed in this study were selected. As indicated earlier, words were selected such that the scale values for the words in each of the ranges (zero-to-20 per cent, 20-to-40 per cent, 40-to-60 per cent, 60-to-80 per cent, and 80-to-100 per cent) fall completely within the boundary values, with median values of 10 per cent, 30 per cent, 50 per cent, 70 per cent, and 90 per cent, respectively. When these median values are compared to the median values obtained for Experimental Condition 13, a higher median value is obtained: i.e., 19 per cent *vs.* 10 per cent, 44 per cent *vs.* 30 per cent, 65 per cent *vs.* 50 per cent, 88 per cent *vs.* 70 per cent, and 98 per cent *vs.* 90 per cent. Also, the ranges of the scale values associated with the median obtained for the five ranges in Experimental Condition 13 (i.e., zero-to-42 per cent, 22-to-74 per cent, 40-to-90 per cent, 54-to-98 per cent, and 84-to-100 per cent) are much wider and are extended towards the high end of the scale, except for the highest of the five ranges—the range of 84-to-100 per cent—which is narrower than in the original sample: i.e., 80-to-100 per cent. This result, the first effect of context to be noted, suggests that when a small subset of stimuli (i.e., 66 words) that varies over a rather narrow range of scale values (i.e., only 20 “units” wide) is removed from a comparatively large population of such stimuli (i.e., 1,000 words) and is judged separately, there is a general expansion and a widening of the absolute scale and of the scale values associated with such stimuli. Experimental Condition 12, a condition in which only three ranges are involved (zero-to-20 per cent, 40-to-60 per cent, and 80-to-100 per cent), yields essentially the same median and range values as does Experimental Condition 13. This result suggests that the important function of Experimental Condition 13 is that it establishes two end ranges and a middle range to which  $S$  may anchor his judgments.

Experimental Conditions 7, 8, 9, 10, and 11, in each of which one of the five ranges was presented alone, show marked differences from Experimental



Condition 13, in which all five ranges were presented together. For ranges zero-to-20 per cent, 20-to-40 per cent, and 40-to-60 per cent, the median scale values are significantly higher in Experimental Conditions 7, 8, and 9 than in Experimental Condition 13; while for ranges 60-to-80 per cent and 80-to-100 per cent, the median scale values are significantly higher in Experimental Condition 13 than they are in Experimental Conditions 10 and 11. This result shows that the effects of context are opposite at the two ends of the scale; that is, when stimuli with scale values distributed evenly over the entire range of the scale are presented for judgment, scale values at the low end of the scale are lowered (35 per cent *vs.* 19 per cent, 60 per cent *vs.* 44 per cent, and 72 per cent *vs.* 65 per cent); while scale values at the high end of the scale are raised (82 per cent *vs.* 88 per cent and 95 per cent *vs.* 98 per cent) in contrast to sets of stimuli with scale values covering only a portion of the entire range.

Experimental Conditions 1, 3, and 5 (in which only two ranges are mixed—40-to-60 per cent and 80-to-100 per cent; 20-to-40 per cent and 60-to-80 per cent, or zero-to-20 per cent and 40-to-60 per cent—rather than all five ranges as in Experimental Condition 13) also show significant and systematic context effects when contrasted with the relevant single ranges of Experimental Conditions 7, 8, 9, 10, and 11. All three experimental conditions (i.e., 1, 3, and 5) show a significant decrease in scale values towards the low end of the scale and a significant increase in scale values towards the high end of the scale when contrasted with the appropriate single ranges of Experimental Conditions 7, 8, 9, 10, and 11: i.e., 72 per cent *vs.* 62 per cent and 95 per cent *vs.* 98 per cent, 60 per cent *vs.* 54 per cent and 82 per cent *vs.* 91 per cent, and 35 per cent *vs.* 20 per cent and 72 per cent *vs.* 76 per cent.

Additional evidence demonstrating the strong effects of context may be found by examining the median scale values for the range 40-to-60 per cent in Experimental Conditions 1, 5, 9, 12, and 13, each of which includes this particular range but embedded in a different context in each of the five different experimental conditions. For Experimental Conditions 12 and 13, the median scale values for the range 40-to-60 per cent are 63 per cent and 65 per cent respectively, not significantly different from each other. For Experimental Condition 9, however, the median scale value increases to 72 per cent, a value significantly greater than the values of 63 per cent and 65 per cent. In Experimental Condition 1, the median scale value of 62 per cent is significantly less than the median value of 72 per cent found in Experimental Condition 9; while in Experimental Condition 5, the median scale value of 76 per cent is significantly greater than the median scale value of

72 per cent found in Experimental Condition 9. Thus, the specific median scale value found for the range 40-to-60 per cent varies directly as a function of the embedding context (or lack thereof as in Experimental Condition 9) and shows systematic and significant increases or decreases as a function of the scale values of the surrounding or embedding stimuli.

## 2. Reinforcement

Experimental Conditions 2, 4, and 6 introduce the operation of reinforcement into the judgmental task; so a trend analysis was made to determine if changes in performance occur as a function of this reinforcement operation.<sup>2</sup> It had been expected that as *S* proceeded through the judgmental task (as, for example, Experimental Condition 2), *S* would judge progressively more and more of the 80-to-100 per cent stimuli as belonging to the concept "economic," while he would judge fewer and fewer of the 40-to-60 per cent stimuli as belonging to that concept. Table 2 shows that, for Experimental Condition 2, the mean number of words judged as belonging is 16.22 words for the 40-to-60 per cent range, and successive 44-word blocks within that range show a systematic and significant drop ( $p < .001$ , as evaluated with the Friedman two-way analysis-of-variance test) from 7.26 words to 4.78 words to 4.18 words. For the 80-to-100 per cent range, the mean number of words judged as belonging is 59.46 words, and successive 44-word blocks show no systematic differences (i.e., 20.10, 19.70, and 19.66 are the values obtained for the three successive word blocks). These results indicate that the reinforcement operation leads to a significant decrease in the number of stimuli judged as belonging to the concept "economic," for the 40-to-60 per cent range; but it has no effect on increasing the number of stimuli judged as belonging to the concept for the 80-to-100 per cent range.

Comparison of the results obtained in Experimental Condition 2 with those obtained in Experimental Condition 1 shows a marked and significant drop (40.36 *vs.* 16.22,  $p < .001$ ) for the 40-to-60 per cent range. However, comparison of the 80-to-100 per cent range also shows a significant drop (98 per cent *vs.* 93 per cent). It had been predicted that the reinforcement operation in the 80-to-100 per cent range, in all probability, would increase the median scale value to 100 per cent, as it was at 98 per cent without reinforcement. This anomalous result suggests<sup>3</sup> that reinforcement in this situation, rather than increasing *S*'s confidence for making a "yes" judgment on the 80-to-100 per cent stimuli, makes *S* more cautious than he is in the complete absence of

<sup>2</sup> It should be noted that trend analyses for successive 44-word blocks were made for all ranges in all 13 experimental conditions, and only the 40-to-60 per cent range in Experimental Condition 2 and the zero-to-20 per cent range in Experimental Condition 6 yield significant shifts in performance over successive 44-word blocks.

reinforcement. The overall result of the reinforcement operation is a shift downward in the scale values of the 40-to-60 per cent stimuli and 80-to-100 per cent stimuli.

A trend examination of the results obtained in Experimental Condition 4 reveals that no changes in judgmental performance occur in successive 44-word blocks; so the reinforcement schedule operates differently than it did in Experimental Condition 2 in which changes in performance occur in successive 44-word blocks in the 40-to-60 per cent range. However, comparison of the results obtained in Experimental Condition 4 with those obtained in Experimental Condition 3 again shows marked and significant drops in median scale values for both the 20-to-40 per cent range and the 60-to-80 per cent range (54 per cent *vs.* 28 per cent and 91 per cent *vs.* 88 per cent). Again as expected, there is a drop in scale values for the 20-to-40 per cent range as a function of reinforcement; but unexpectedly, the scale values in the 60-to-80 per cent range (rather than increasing towards a median scale value of 100 per cent) show a slight but significant drop (91 per cent *vs.* 88 per cent).

A trend examination of the results obtained in Experimental Condition 6 shows, for the zero-to-20 per cent range, a systematic and significant drop ( $p < .01$ ) in successive 44-word blocks in the number of words judged as belonging to the concept (i.e., 5.66 to 4.18 to 3.72), but no systematic trend is found in the 40-to-60 per cent range. However, comparison of the results obtained in Experimental Condition 6 with those obtained in Experimental Condition 5 shows a drop in the number of words as belonging in the zero-to-20 per cent range (14.58 *vs.* 13.26), but the difference is not significant; while, for the first time, there is a significant increase (76 per cent *vs.* 86 per cent) for the 40-to-60 per cent range.

The results obtained in Experimental Conditions 2, 4, and 6 suggest that the effects of reinforcement operations are complex and are far from uniform over the entire range of the scale. Trend effects are found, for example, in only two of the six possible ranges in which such effects had been expected. Also, in two of the three ranges in which reinforcement had been expected to produce more "yes" judgments, it produced fewer "yes" judgments.

In the evaluation of the entire set of results obtained in this study, as they relate to context and reinforcement operations, there is little available at present in the way of a theoretical model that will integrate the rather diverse findings—particularly the seemingly complex interaction of reinforcement operations and subrange scale values. It seems more appropriate, therefore, to suggest some further lines of inquiry that should provide a broad empirical base on which future models might be constructed.

The following lines of investigation would seem to be most relevant: (a) a systematic study of the effects of the distance in scale units that separate two ranges; (b) an examination of the effects of varying proportions of stimuli from two or more ranges; (c) an investigation of the effects of reinforcement applied in reverse: i.e., reinforcing low scale values as correct and reinforcing high scale values as wrong; and (d) a study of the effects of mixing wider ranges (e.g., 30 or 40 scale units) or mixing narrower ranges (e.g., five or 10 scale units).

## F. SUMMARY

650 Ss were tested individually (50 Ss in each of 13 different experimental conditions) in a study of the effects of context and reinforcement on the judgment of verbal stimuli. Ss were presented with a series of words, each of which was to be judged by each S for inclusion in the concept "economic." The words had been "scaled" previously with respect to inclusion in the concept, and the various experimental conditions provided varying mixes of scale-range values and reinforcement schedules. The results show significant but complex effects of context and reinforcement on judgmental behavior. Future lines of inquiry are indicated.

## REFERENCES

1. MAYZNER, M. S., & TRESSELT, M. E. Shifts in connotative meaning of words as a function of previous restrictive experience. *J. Exper. Psychol.*, 1958, **55**, 200-205.
2. ———. Shifts in connotative meaning of words as a function of varying amounts of previous restrictive experience. *J. of Psychol.*, 1959, **47**, 107-116.
3. ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, No. 261.
4. SIEGEL, S. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1956.
5. TRESSELT, M. E. The influence of amount of practice upon the formation of a scale of judgment. *J. Exper. Psychol.*, 1947, **37**, 251-260.
6. TRESSELT, M. E., & MAYZNER, M. S. Consistency of judgments in categorizing verbal material. *Psychol. Rep.*, 1958, **4**, 415-421.
7. ———. A further study in the consistency of judgments in categorizing verbal material. *J. of Psychol.*, 1959, **47**, 337-341.
8. ———. Scale values for 1000 words judged for inclusion in the concept "economic." *Psychol. Newsl.*, 1959, **10**, 350-358.

*Department of Industrial Engineering  
and Operations Research<sup>a</sup>  
New York University  
University Heights  
Bronx 53, New York*



## CRITICAL FLICKER FREQUENCY FOR HIGH PULSE-TO-CYCLE FRACTIONS AND HIGH INTENSITY\*<sup>1</sup>

*Department of Psychology, Michigan State University*

JANE E. RANNEY AND S. HOWARD BARTLEY

### A. PROBLEM

Bartley and co-workers have presented data in several articles (1, 2, 3, 4) describing critical flicker frequency (CFF) as a function of pulse-to-cycle fraction (PCF) under intensities varying from 1.28 c/ft<sup>2</sup> to 1164 c/ft<sup>2</sup>. Under the highest intensity (1164 c/ft<sup>2</sup>) the CFF for PCF 7/8 was higher than that for PCF 3/4. Bartley and Nelson state that this result "... may be artifactual" (2, p. 6). This reversal constituted a peculiar irregularity in the curve, similar to the one at the low PCF end of the same curve which irregularity was accounted for. It was in line with the finding that two PCFs would produce the same CFF under the same intensity conditions, whereas traditionally every PCF was supposed to function differently.

If the irregularity in the high PCF end of the curves were to be established, then it would indicate that two very high PCFs could function the same. If so, one possibly would have to modify the supposition that the off-effect plays a role in determining flicker under high PCF conditions.

Due to the need to check on what has proved so far to be a serviceable model, the present new observations were made.

### B. METHOD

#### 1. *Observers*

Five observers participated, three of whom were sophisticated observers who served in the original research. The other two observers had about three hours of practice in making CFF judgments prior to this experiment.

#### 2. *Apparatus*

A General Electric ribbon filament lamp was mounted behind opal glass, with a single-open-sector 25.6-cm plastic episcotister disc rotating in front of

\* Received in the Editorial Office on June 5, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This investigation was supported in part by a Public Health Service fellowship 1 Fl MH-22,046-01 from the national Institute of Mental Health and in part by Research Grant NSF-G-19485 from the National Science Foundation.



it. Intensity of the target was 3000 c/ft<sup>2</sup>, and was reduced to 1200 and 1890 c/ft<sup>2</sup> by 0.4 and 0.2 Kodak Wratten neutral density filters respectively. Pulse rate was controlled by a Variac. A Weston voltage generator connected to the episcotister shaft and voltmeter calibrated in pulses per second provided the CFF readings.

A "bite board" maintained the observer's head position 76 cm from the target which was two cm by one cm and thus subtended 1° 30' by 45' visual angle. A reduction screen with a .8-cm circular opening was placed 12 cm in front of him.

### 3. Procedure

The observer was seated comfortably and the five PCFs (1/2, 5/8, 3/4, 7/8, 29/30) were presented in random order. For each PCF, the observer was instructed to adjust the pulse rate with the Variac, increasing it until the target (rather than the surround) no longer appeared to flicker, then increasing the rate beyond this point and decreasing it until flicker was again noticeable. Twelve alternately ascending and descending monocular observations were taken under each condition, but the first two were discarded as practice trials. Mean CFFs and standard deviations were computed for each observer.

As a matter of convenience, the conditions in this experiment differed slightly from those of the original research: (a) The target subtended 1° 30' in this experiment rather than 1° 4' as in the original investigation. In both cases the stimulation was foveal. (b) The intensity was 1200 c/ft<sup>2</sup> rather than 1164 c/ft<sup>2</sup>. However these differences are slight and would not be expected to have significant effect on the shape of the CFF curve. In addition, to confirm results at this intensity, three observers made observations at an intensity of 1890 c/ft<sup>2</sup>.

A third difference was also involved. Instead of investigating only one or two PCF conditions in a single experimental session, data for all PCF conditions were gathered in each session. It was thought possible that the irregularity in the original curve may have arisen, at least in part, from the fact that not all PCFs were sampled in each session. It is well known that observers generally vary more from day to day than from moment to moment. Such a variation in this experiment naturally would have an effect on the general shape of the curve in question. Hence we gathered data for all PCFs at each session.

### C. RESULTS AND DISCUSSION

Mean CFFs and standard deviations for each observer are shown in Figure 1A. From this graph it is obvious that there is no increase in CFF for

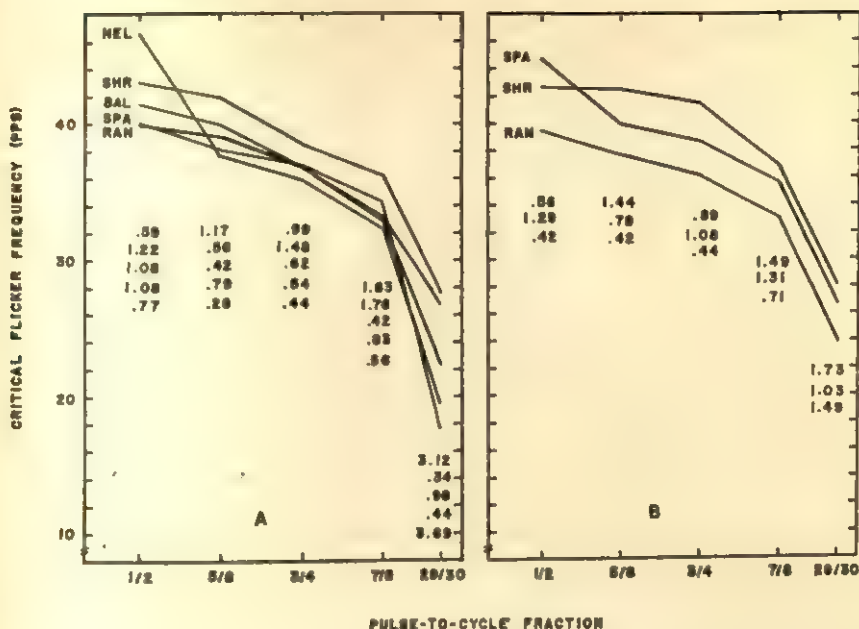


FIGURE 1

CRITICAL FLICKER FREQUENCY AS A FUNCTION OF PULSE-TO-CYCLE FRACTION

A presents mean CFFs for five observers as a function of PCF for intensity of 1200 c/ft<sup>2</sup>. B presents mean CFFs for three observers as a function of PCF for intensity of 1890 c/ft<sup>2</sup>. Standard deviations are listed in the same order as the points plotted.

PCF 7/8 over that of PCF 3/4, but rather a consistent decrease for all five observers. The same is true for the higher intensity, 1890 c/ft<sup>2</sup> (Figure 1B). Therefore we conclude that the irregularity in the original data was an artifact.

## REFERENCES

1. BARTLEY, S. H., & NELSON, T. M. Equivalence of various pulse-to-cycle fractions in producing critical flicker frequency. *J. Opt. Soc. Amer.*, 1960, **50**, 241-244.
2. ———. Some relations between pulse-to-cycle fraction and critical flicker frequency. *Percept. & Motor Skills*, 1960, **10**, 3-8.
3. ———. A further study of pulse-to-cycle fraction and critical flicker frequency. A decisive theoretical test. *J. Opt. Soc. Amer.*, 1961, **51**, 41-45.
4. NELSON, T. M., BARTLEY, S. H., & DEHARDT, D. A comparison of variability of three sorts of observers in a sensory experiment. *J. of Psychol.*, 1960, **49**, 3-11.

Department of Psychology  
Michigan State University  
East Lansing, Michigan

## CFF FOR SHORT TRAINS OF PHOTIC STIMULATION HAVING VARIOUS TEMPORAL DISTRIBUTIONS AND SEPARATIONS\*<sup>1</sup>

*Department of Psychology, Michigan State University*

THOMAS M. NELSON, S. HOWARD BARTLEY, AND EARL S. HARPER

### A. INTRODUCTION

The time separation between individual pulses of a photic train does not have a constant effect upon sensory end result. Separations sufficient to produce vigorous flicker, when long sequences of photic stimulation are employed, become ineffective as components of short trains. Within a broad range of intensities, rate of photic intermittency must increase as the number of pulses in the train increases, if fusion is to be maintained (1, 4). The function describing the increase is negatively accelerated: at first, increasing the number of pulses from the lower limit of two produces very large changes in CFF (critical flicker frequency), but before a train of 30 pulses is employed an asymptotic value is reached (4).

Limiting the number of pulses in a train necessitates control of the time interval separating successive trains. There is certain to be some effect from systematic variation of that interval. Progressive decreases in the interval between short trains finally produce a condition in which the interval equals the separation between pulses in the train. This increases CFF.

It should not be thought that short-train sequences of stimulation must always involve a separating interval possessing a uniform characteristic. Forsyth and Brown (2) and Forsyth and Chapanis (3) have interlaced frequencies so as to form complex repetitive patterns, but the use of such stimulus patterns produces events not ideally suited to neurophysiological analysis.

Response to stimulation by short trains has general empirical and theoretical importance. The use of such inputs provides data which will begin to bridge the gap existing between stimulus conditions employing single inputs and paired inputs on the one hand and those involving extended periods of stimulus intermittency on the other. The research to be reported here ex-

\* Received in the Editorial Office on June 5, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The present paper is a part of the work being done under Research Grant NSF-G-19485 from the National Science Foundation.

tends the range of conditions hitherto employed in CFF investigation by using trains ranging from three to 23 pulses in conjunction with (a) pulse-to-cycle fractions (PCF) varying from .17 to .83 and (b) separating intervals of no stimulation representing 1.17 cycles to 21.83 cycles.

### B. METHOD

A disk (episcotister) four feet in diameter produced stimulus intermittency. Rate of delivering the pulses was varied by changing the rotation rate of the disk and was measured by a Weston Tachometer.

Five disks were used. Each disc was sectored so as to produce 24 identical stimulus cycles each revolution. Each cycle on the disk having the largest PCF involved an open sector five times the value of the closed sector. The PCF of each cycle produced by this disk was therefore .83. The remaining disks were cut so as to produce PCFs of .67, .60, .33, and .17.

Pulse trains having 3, 5, 7, 9, 11, 19, and 23 unit sequences of photic repetition were employed. These were separated by varying the number of closed sectors intervening between pulse trains. Since the number of total sectors on the disk peripheries was always 24, possibilities of pulse separation were lim-

#### SAMPLE PHOTIC PULSE SEQUENCES

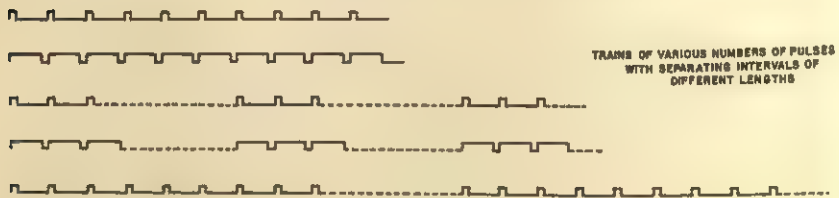


FIGURE 1

#### STIMULUS CYCLING AS USED IN THE PRESENT EXPERIMENT

In the upper two rows, pulses are equally separated. The lower rows show sequences consisting of component trains varying in number of photic pulses and in train separations. Absolute values of all components, of course, shift with rate of cycling.

ited. The combinations sampled were the following: three pulse trains separated by 1, 3, 5, 9, and 21 closed sectors; five pulse trains separated by 1, 3, 7, and 19 closed sectors; seven pulse trains separated by 1, 5, and 17 closed sectors; nine pulse trains separated by 3 and 15 closed sectors; 11 pulse trains separated by 1 and 13 closed sectors; a 19-pulse train separated by five closed sectors; and a 23-pulse train with one closed sector. A sequence involving 10 seconds of uninterrupted intermittent stimulation was also used.

Figure 1 is a schema of several of the input conditions used. Notice two things. First, the time between successive trains is not fixed but decreases as stimulation rate increases. Second, the absolute separation between the last pulse in one train and the first pulse in the succeeding train, at any given rate of stimulation, is less for the larger PCFs than for the smaller. One must add the time occupied by the closed portion of the photic cycle to the time occupied by cycles completely closed. Thus, at a stimulation rate of 10 cps, PCF .17 trains separated by three closed sectors were separated by 383 milliseconds, while the three-closed-sector separation between PCF .33, .50, .67, and .83 resulted in 367, 350, 333, and 317 milliseconds respectively.

The target was an opal glass surface passing radiation from a GE incandescent "artificial-daylight" filament lamp. At the distance used, the target subtended  $2^{\circ} 5'$  in the vertical and  $45'$  in the horizontal. The disk swept across the horizontal dimensions. Onset and termination time represented 29 per cent of the total cycle. Thus the photic periods of smallest and null periods of the largest PCFs were greatly tapered.

The intensity of the unfiltered target (intensity 0) was  $5340 \text{ c/ft}^2$ . For the other targets, intensity was reduced in logarithmic steps by means of Wratten neutral density filters. Four log reductions were used resulting in intensities of 534, 53.4, 5.34, and  $.53 \text{ c/ft}^2$  and were designated as intensities 1, 2, 3, 4.

### C. PROCEDURE

All observations were made monocularly from a chin rest. Two male students trained in making flicker-fusion discrimination served as *Os*. CFF was established by the psychophysical "method of limits." An assistant adjusted the speed of the disk in alternately ascending and descending directions until *O* reported flicker totally abolished or reintroduced.

Successive presentation of trains involved separations between pulse trains considerably longer than the null stimulus periods between pulses making up the train. At slow rates of intermittency, the appearance is therefore of a train of flicker periodically interrupted. The *Os* fused only the component trains. Above CFF, a steady target field appears to be interrupted periodically. In this respect the procedure is in contrast to that used by Forsyth and Brown (2) and Forsyth and Chapanis (3), whose *Os* fused the total of a complex temporal distribution.

### D. RESULTS AND DISCUSSION

The effects obtained from increasing the number of pulses in the train is given in the first and second columns of Figure 2 for PCF .33. Both *Os* are



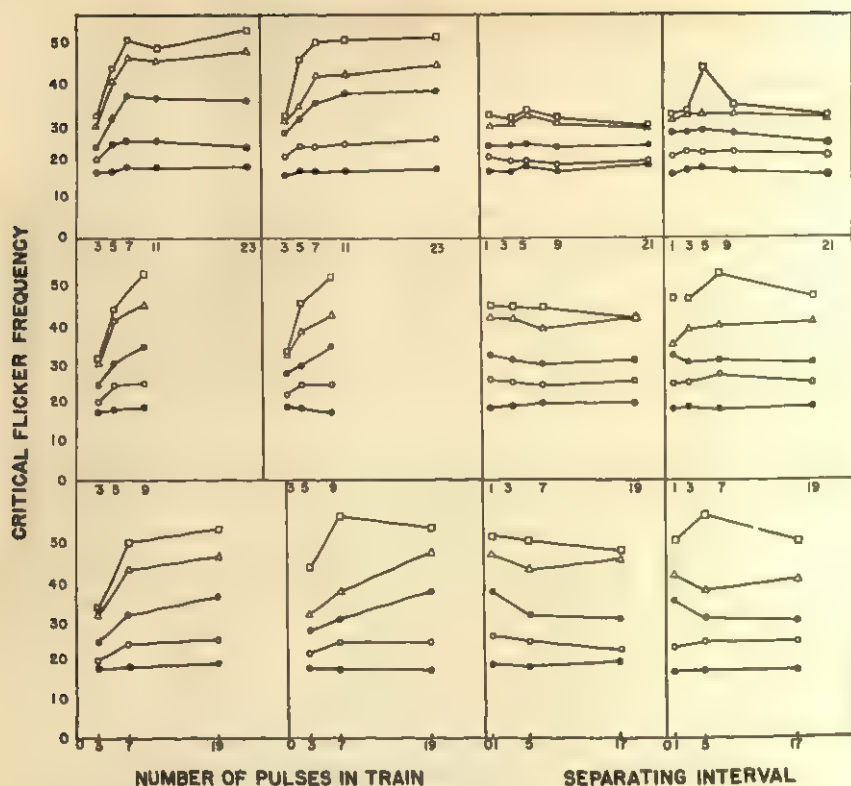


FIGURE 2

CRITICAL FLICKER FREQUENCY CURVES FOR OBSERVERS S. H. AND F. B.

The first and third columns contain curves for observer S. H. The second and fourth columns contain curves for observer F. B. In the first and second columns, each pair of panels is for a constant separating interval. In the first pair, the separating interval is 1; in the second pair, the separating interval is 3; and in the third pair, it is 5. The curves are ordered according to level of intensity, with the open square representing the highest level and the filled circles a condition 4 log units less.

In the third and fourth columns, each pair represents the results from varying separating intervals while holding number of repetitive stimulations constant. The upper pairs are for a 3-pulse train, the second pair for a 5-pulse train, and the lowest pair for a 7-pulse train. Intensity level is ordered as in Columns 1 and 2.

represented. Each family of curves depicts changes at the five levels of intensity, with a constant separating interval. Excepting for the lowest intensity conditions, all curves are positively accelerated reaching asymptote somewhere between nine and 11 units. Complete data for all PCFs are given in Tables 1 to 5. Except for the lowest intensity level, an increase in rate is required throughout to produce CFF as number of pulses in the train is increased.

TABLE 1  
CRITICAL FLICKER FREQUENCY FOR PULSE-CYCLE FREQUENCY 1/3 AT FIVE  
INTENSITY LEVELS

Number of pulses in train— separating interval	Observer S. H.					Observer F. B.				
	0	1	2	3	4	0	1	2	3	4
3 - 1	32.83	29.67	24.50	21.67	18.00	32.17	30.67	27.50	21.50	17.17
3 - 3	31.33	30.00	24.67	20.33	17.67	33.00	32.33	27.67	22.50	18.83
3 - 5	33.83	32.17	25.00	20.67	18.17	44.17	32.00	28.00	22.17	18.17
3 - 9	31.17	30.00	24.00	19.50	17.67	34.33	32.50	27.00	22.50	17.50
3 - 21	29.00	28.83	24.33	20.50	19.00	32.00	31.33	25.00	22.00	17.50
5 - 1	44.83	41.67	32.00	25.50	18.33	46.17	34.17	31.83	24.00	17.67
5 - 3	44.33	41.50	30.67	24.83	18.50	45.50	38.17	29.33	24.50	18.17
5 - 7	44.00	38.92	29.50	24.08	19.17	52.33	39.17	30.33	26.67	17.83
5 - 19	41.17	41.67	30.33	25.00	19.33	46.83	40.17	29.33	24.33	18.33
7 - 1	52.33	47.67	38.00	26.17	19.50	51.17	42.00	35.67	23.83	17.33
7 - 5	51.00	43.83	32.00	24.67	18.66	57.83	38.00	31.00	25.00	17.67
7 - 17	48.00	46.00	30.83	23.17	19.83	51.17	41.00	30.50	25.00	17.50
9 - 3	53.33	45.50	34.33	25.17	19.00	52.33	42.50	34.67	24.33	17.33
9 - 15	53.33	46.50	32.67	24.67	20.17	55.67	41.50	32.50	26.00	17.17
11 - 1	54.00	47.33	37.00	26.00	19.00	51.67	43.17	37.83	24.67	17.33
11 - 13	49.67	47.17	33.67	25.17	19.17	53.00	42.50	33.67	25.83	18.33
19 - 5	54.67	47.17	37.00	25.17	19.33	54.67	48.17	38.17	25.17	17.83
23 - 1	54.33	49.17	36.17	24.33	19.00	52.17	44.83	38.00	25.33	18.00
10 sec.	51.83	50.00	36.17	27.00	19.00	53.50	47.00	37.83	26.30	18.50

TABLE 2  
CRITICAL FLICKER FREQUENCY FOR PULSE-CYCLE FREQUENCY 1/6 AT  
FIVE INTENSITY LEVELS (OBSERVER S. H.)

Number of pulses in train— separating interval	Intensity level			
	0	1	2	3
3 - 1	33.67	32.33	27.17	22.83
3 - 3	36.83	37.00	27.50	23.33
3 - 5	36.17	31.50	28.83	21.50
3 - 9	36.33	37.50	29.00	22.33
3 - 21	35.50	35.17	26.17	21.50
5 - 1	48.33	42.67	34.17	23.83
5 - 3	46.83	35.67	30.83	23.67
5 - 7	47.50	36.33	32.00	24.17
5 - 19	48.17	42.17	30.50	23.67
7 - 1	54.83	41.33	34.50	23.33
7 - 5	51.50	41.17	34.83	23.50
7 - 17	56.17	42.83	32.17	24.50
9 - 3	54.83	43.00	34.67	23.33
9 - 15	57.83	42.00	33.33	24.50
11 - 1	52.67	43.33	37.00	23.17
11 - 13	55.00	43.00	33.50	24.17
19 - 5	57.17	45.83	36.33	24.00
23 - 1	56.17	46.33	37.33	23.00
10 sec.	58.17	48.00	34.67	23.17

TABLE 3  
CRITICAL FLICKER FREQUENCY FOR PULSE-CYCLE FREQUENCY 1/2 AT  
FIVE INTENSITY LEVELS (OBSERVER S. H.)

Number of pulses in train— separating interval	Intensity level				
	0	1	2	3	4
3 - 1	35.00	34.33	30.50	23.00	17.33
3 - 3	33.83	33.67	27.00	20.83	18.33
3 - 5	35.00	33.50	26.83	20.00	17.83
3 - 9	37.17	34.67	25.17	19.00	17.83
3 - 21	41.00	32.67	25.17	21.33	16.83
5 - 1	49.67	44.00	34.17	25.00	17.83
5 - 3	45.17	41.33	30.00	22.67	18.17
5 - 7	50.00	43.17	28.00	23.67	17.67
5 - 19	45.33	41.50	28.33	23.83	17.83
7 - 1	53.50	47.07	37.33	26.33	19.50
7 - 5	54.83	46.00	31.17	22.83	19.50
7 - 17	50.50	43.67	31.50	25.17	17.67
9 - 3	52.33	44.33	34.00	24.83	18.17
9 - 15	53.33	45.00	31.83	25.17	18.00
11 - 1	56.67	48.17	36.17	25.67	18.50
11 - 13	52.33	44.83	30.50	27.33	18.00
19 - 5	60.67	47.83	35.67	23.17	19.83
23 - 1	58.00	46.83	38.17	23.67	18.83
10 sec.	59.17	50.83	36.33	25.67	17.83

TABLE 4  
CRITICAL FLICKER FREQUENCY FOR PULSE-CYCLE FREQUENCY 2/3 AT FIVE  
INTENSITY LEVELS

Number of pulses in train— separating interval	Observer S. H.					Observer F. B.				
	0	1	2	3	4	0	1	2	3	4
3 - 1	30.50	28.92	25.33	22.66	20.55	32.75	26.25	21.08	18.42	14.92
3 - 3	27.75	28.92	23.58	19.83	17.33	30.75	27.75	23.67	19.08	16.42
3 - 5	30.33	31.92	25.08	21.50	19.33	29.67	26.83	23.00	19.75	17.00
3 - 9	30.92	27.50	23.67	19.20	16.50	29.00	26.42	22.25	18.58	16.50
3 - 21	24.50	25.67	24.00	18.92	18.25	26.67	27.42	23.58	18.92	16.17
5 - 1	43.75	36.08	31.17	23.25	18.67	37.42	32.83	25.67	19.67	15.75
5 - 3	38.75	34.00	28.25	21.42	18.17	39.25	33.17	26.75	21.50	16.83
5 - 7	38.67	38.92	26.42	20.75	17.28	39.67	34.83	27.58	22.58	18.08
5 - 19	34.58	35.42	27.83	22.42	19.83	38.17	35.33	26.33	23.08	17.92
7 - 1	46.25	40.92	33.17	24.33	15.58	44.83	39.42	29.42	22.50	16.92
7 - 5	45.17	42.00	28.75	23.92	18.25	43.75	38.92	27.92	22.08	17.83
7 - 17	45.42	44.83	31.83	22.33	18.92	39.67	38.08	28.33	23.42	18.58
9 - 3	50.08	42.42	34.00	22.83	17.83	43.75	40.17	30.25	23.33	17.00
9 - 15	52.67	46.83	29.83	22.67	18.82	43.17	39.50	29.00	23.58	18.25
11 - 1	46.42	48.58	36.75	24.58	19.08	47.75	42.75	32.33	22.33	17.67
11 - 13	45.42	46.48	33.42	23.33	20.25	43.08	38.00	30.08	23.08	17.75
19 - 5	48.25	46.50	35.66	24.58	19.00	47.58	43.17	32.17	23.17	18.08
23 - 1	50.83	50.00	38.20	27.42	22.17	47.83	44.58	33.17	22.67	17.42
10 sec.	52.50	48.70	35.20	25.20	19.00	45.25	42.50	31.25	21.00	15.50

TABLE 5  
CRITICAL FLICKER FREQUENCY FOR PULSE-CYCLE FREQUENCY 5/6 AT  
FIVE INTENSITY LEVELS (OBSERVER S. H.)

Number of pulses in train— separating interval	Intensity level				
	0	1	2	3	4
3 - 1	27.50	26.83	23.17	19.33	16.83
3 - 3	27.83	24.17	24.17	19.00	16.33
3 - 5	26.83	24.83	22.33	17.17	15.33
3 - 9	27.17	31.00	22.83	15.67	16.00
3 - 21	24.00	23.83	20.00	16.33	15.50
5 - 1	40.17	37.67	29.50	20.50	15.83
5 - 3	36.50	32.50	24.00	20.00	15.50
5 - 7	31.83	37.17	23.67	18.33	15.83
5 - 19	32.50	27.67	22.00	18.00	16.33
7 - 1	44.33	40.33	31.33	19.50	16.67
7 - 5	39.17	41.00	25.83	20.83	15.83
7 - 17	36.00	35.83	24.17	20.00	17.17
9 - 3	41.67	38.67	28.17	20.83	16.17
9 - 15	37.67	40.17	25.00	18.17	16.00
11 - 1	45.50	42.17	33.33	22.33	16.17
11 - 13	42.00	40.00	26.00	20.50	17.50
19 - 5	44.17	42.83	30.17	18.67	15.17
23 - 1	45.83	44.00	33.33	23.33	15.50
10 sec.	47.00	46.17	32.67	20.83	15.67

Tables 1 to 5 also show that the greatest range of CFF values occurred with PCF .50. The decline in the effectiveness of CFFs is asymmetric: larger PCFs fall off more rapidly than the smaller. Not only is the asymptote more rapidly reached with the larger PCFs, but these also start at a lower value. Flicker appears much weaker at the extreme PCFs.

Separating interval has virtually no effect upon CFF. This is clearly shown in the graphs in the two right hand columns of Figure 2 as well as in Table 3. It is to be expected, however, that manipulation of separations would have been effective with separation magnitudes smaller than we used.

#### E. CONCLUSIONS

The sensory effects of varying PCF, intensity, and separating interval are given for PCF .33 as a series of three-dimensional graphs of Figure 3. In each case, there is an increase in CFF as pulse train lengthens, but there is no change from varying separating interval. The depiction, of course, is incomplete since a further progressive decrease in separating interval would have the effect of raising the front edge of the upper models to the asymptotic CFF values of the longest pulse train. The models also omit from consideration manipulations in which the separating interval is at the photic level of the pulses instead of being a period of no stimulation.

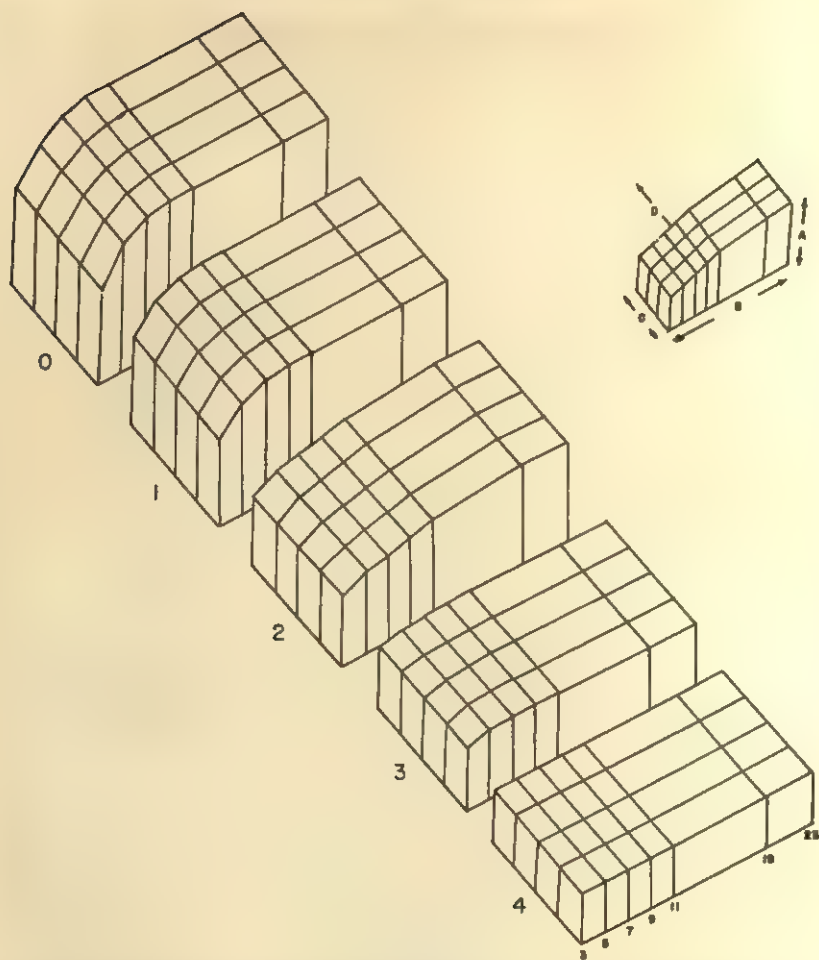


FIGURE 3

THREE-DIMENSIONAL REPRESENTATIONS OF THE RELATION BETWEEN CFF (A),  
NUMBER OF PULSES IN THE TRAIN (B), SEPARATING INTERVAL (C), AND  
INTENSITY (D) FOR THE PCF 1/3 CONDITION (OBSERVER S. H.)

The upper model (0) represents the highest intensity used and the lower model (4) represents a 4-log-unit reduction in intensity.

#### F. SUMMARY

Study was made of the effects upon CFF of combinations of five pulse-to-cycle fractions (PCFs), five levels of intensity, eight lengths of pulse trains, and 11 separation intervals between trains.



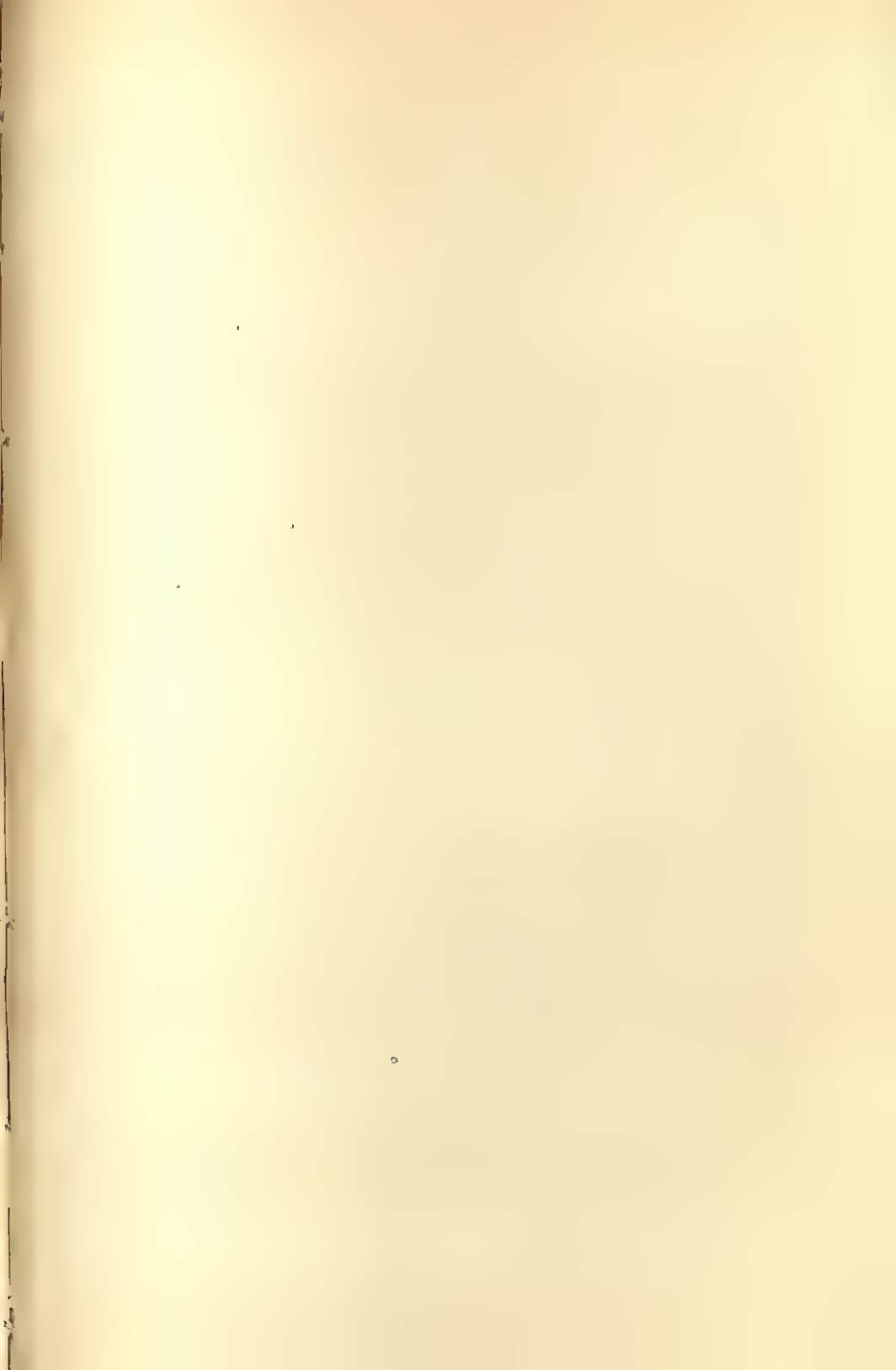
It was discovered that (a) CFF increases as the number of pulses in the train increases except at the lowest levels of intensity, (b) intensity varies CFF magnitude and range, and (c) PCF varies CFF magnitude and range when length of pulse train is varied. Variation of separating intervals produced no effect in this experiment although it is clear that such would have happened had further manipulations been made.

## REFERENCES

1. BARTLEY, S. H., NELSON, T. M., & RANNEY, J. E. The sensory parallel of the reorganization period in the cortical response in intermittent retinal stimulation. *J. of Psychol.*, 1961, **52**, 137-147.
2. FORSYTH, D. M., & BROWN, C. R. Flicker contours for intermittent photic stimuli of alternating durations. *J. Opt. Soc. Amer.*, 1959, **49**, 760-763.
3. FORSYTH, D. M., & CHAPANIS, A. Counting repeated light flashes as a function of their number, their rate of presentation, and retinal location stimulated. *J. Exper. Psychol.*, 1958, **56**, 385-391.
4. NELSON, T. M., & BARTLEY, S. H. The Talbot-Plateau Law and the brightness of restricted numbers of photic repetitions at CFF. *Vision Res.*, in press.

*Department of Psychology*  
*Michigan State University*  
*East Lansing, Michigan*







## THE EFFECTS OF CYCLING UPON CFF WHEN STIMULUS PURITY IS VARIED\*<sup>1</sup>

*Department of Psychology, Michigan State University*

THOMAS M. NELSON, S. HOWARD BARTLEY, AND FRED C. BLECK

### A. INTRODUCTION

When analysing natural events, one may lose sight of the fact that all exact description involves a large component of formal analysis. There is a tendency to fall into an everyday commonsense way of thinking that does not recognize description by measurement as separate from the thing measured. Exact descriptions are considered natural in themselves. And one comes to think of phenomena in terms of one set of values rather than another, overlooking the simplicity that might be gained and difficulties that might be circumvented by altering the route of theoretical development.

For example, present knowledge suggests that the functioning visual system is more amenable to temporal than to spatial analysis. The effects of stimulation may be described in terms of amplitude, frequency, and form; and may be followed by making recordings of neurophysiological activity at successive locations from periphery to center. One may further assume that the records at every locus describe the net result of facilitation, inhibition, latency, and other characteristics of the particular neural structures involved. In making temporal analysis, one attends to the serial complexity of the propagational mechanism and its manifestation of the principles of nerve excitation (27). The retina is, in contrast, organized into overlapping and interconnected response units. This complexity is altered at the geniculate level, which in turn is unique from the optic cortex. Tissue thought of in purely spatial terms yields no action picture. It would seem that the very pattern of spatial detail reduces the likelihood of "hitting upon" any adequate set of spatial parameters. Adequate analysis of temporal distribution of activity seems more readily accomplished. The difference in temporal descriptions will have more general explanatory value than an equally detailed set of static factors even though the role of spatial factors<sup>2</sup> will ultimately be necessary for understanding many of the details of sensory end result.

\* Received in the Editorial Office on June 8, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The present paper is a part of the work being done under Research Grant NSF-G-19485 from the National Science Foundation.



In spite of this, nowhere has a greater exclusive preference for spatial theory been manifested than in sensory psychology and sensory physiology. To rely solely upon an outlook supposing that different conglomerates of receptor cells are responsive to different ranges of wavelengths, as certain color theories do, is simply to rely upon values obtained from certain measurement scales in preference to others without adequate empirical evidence one way or the other. Measurement systems describe because they provide *a priori* frameworks interpretive of concrete events, but it must be conceded that these constitute theories themselves and are very different from the literal entity to which they refer.

Temporal manipulation of stimulation and temporal concepts have played a restricted role in investigations of color and development of color theory. Traditionally, controversy has centered around the properties of the sense cells that produce color. Virtually all theorists—Young, Helmholtz, Hering, Ladd-Franklin, and Polyak—looked for localized properties that would give color. Time entered incidentally only insofar as it obviously required time to conduct the “message” to the central organ. But time was not a focal concern.

Troland's theory (26) stands as the first comprehensive attempt at temporal theory. He hypothesized that sufficiently complex temporal patterns of discharge could mediate colors. It was Troland's notion that one cone and one retino-cortical path could generate and conduct all possible variations of hue, saturation, and brightness. Fry's development (8) of Troland's theory assigns brightness to the basic frequency of discharges and hue to the distribution (modulation) characteristic of the discharges. Red, yellow, green, and blue are treated as basic modulation forms. Saturation is attributed to amplitude of a given distribution characteristic. Other colors (orange, cyan, lemon) represent intermediate modulations. The chief difficulty with the Troland-Fry formulation, as both Troland and Fry acknowledge, has been that the photoreceptor mechanism producing varying modulations seems to require receptors possessing unusual properties. Fry suggested five different types of exciting substances for each photoreceptor.

More recent neurophysiological investigations have been made into differences in the timing of discharges induced by intermittent part-spectrum stimulation. These seem to offer an avenue for empirical but nonspatial characterization of sensory activity. Such an approach to color theory seems particularly promising in that it allows retention of useful structural constructs; hence theory may accommodate many of the basic facts of color vision, such as spectral visibility, mixture, complementary colors, and color blindness. On the other hand, it provides concepts seemingly necessary if one is to accommodate color phenomena associated with changes in timing of

input. Since recent neurophysiological investigations have shown that hue, saturation, and brightness changes can be induced by intermittent presentation of part- and full-spectrum stimulation of the eye (2, 3, 5, 18, 20, 21), temporal considerations become directly relevant in color theory.

### B. GENERAL PROBLEM

It is possible that CFF varies with different compositions of wavelength. On the negative side, one may imagine that if each wavelength excites an equal number of receptors and hence engages an equal number of peripheral channels (i.e., if luminances are equal), each will thereby produce the same CFF. Such theorists imagine the presence of anatomically or functionally separate color systems and assume that differences in photochemical process lead to activation of one color system as opposed to another. Action properties of these systems are treated as being the same over time when stimulated to the same degree and by the same temporal pattern.

If one imagines unique patterns of discharge associated with unique colors, it is almost mandatory that CFF vary with wavelength in some way when luminosities are equal. The reason for this rests both upon the general temporal hypothesis (already discussed) and various central nervous facts that have been uncovered from subjecting the visual system of the human and other mammals to part-spectrum stimulation.

It has been known for a long time (11, 22, 24) that there is variability in the time constants for each hue at a given input level. Hues are ordered in latency of establishment from red to blue. Facts more recently found in studies of the optic pathway correlate neurophysiological processes with these differentials of time. Chang (4) has reported faster conduction rates for the longer wavelengths of the visible spectrum. Lennox (16) discovered that "fast" fibers respond more actively to the longest wavelengths than to the shortest wavelengths. The work of Madsen and Lennox (17) shows that cortical latency to "blue" is greater at the center than it is for "green," "yellow," or "red" when input intensities are adjusted to produce equal retinal latencies. Lennox (15) also discovered that response latencies at the lateral geniculate are shortest for long-wavelength stimuli, longer to mid-wavelengths, and longest to short wavelengths. Discharge components are apparently modified by wavelength (16). "On" fibers are more likely to be sensitive to short wavelengths and tend to belong to slower conduction groups. Moreover these fibers are generally longer in latency, productive of fewer spikes, and of a lower discharge frequency than the other types of fibers. "On-off" are opposite in character.

Spontaneous activity is present, whereas it is almost entirely absent in the "on."<sup>2</sup>

Unfortunately, no clear-cut relationships have been established between spectral distribution and CFF. Crozier and Wolf (6) and Hecht and Schlaer (9) report no effects due simply to shifts in part-spectrum stimulation. On the other hand Ives (11), Allen (1), Landis (14), and Nelson, Bartley, and Bleck (19) report evidence in the affirmative. Part of Landis' endeavor consisted in reanalyzing the data of Hecht and Schlaer. In one experiment (19), CFFs were ordered in magnitude from red to blue when fields were equated in luminosity by the method of flicker.

Only one of the studies reported data relating spectral purity to CFF (19). Purity of stimulation was varied by mixing part-spectrum with full-spectrum stimulation. The data indicate effects in relation to the composition of the mixture. When the purity of a part-spectrum input is progressively reduced by greater admixtures of a full-spectrum component, CFF shifts toward the value of the latter.

In other studies, purity has been manipulated to discover its effects upon CCFF (critical color flicker frequency). Determinations were made both with alternating part- and full-spectrum inputs (7, 10, 11, 12, 25) and alternating different part-spectrum inputs (28). All indicate inhomogeneities of CCFF result from the mixing of wavelengths.

Related to this is the fact that photometric values resulting from the method of flicker are known to be affected by the sector ratios employed in flicker photometry (7, 13). Ferree and Rand (7) argue that brightness differences so produced have a temporal explanation. They propose that such effects are due to differential rates of rise of color discrimination. There is weight to this suggestion. The neurophysiological facts revealed lead one to expect general differences in the slope of CFF purity curves, with variations in PCF.

### C. SPECIFIC PROBLEM

The present study reports a further probe into the CFF-wavelength relationships. The notion that flicker depends upon the presence of irregularities in density of massed cortical activity (a theoretical paper in preparation) leads to two hypotheses. First, CFF ought to vary with PCF in the same way for all spectral distributions. Pulse duration increases linearly with PCF. As one approaches the value of steady stimulation (PCF, 1.00) the likelihood of faster

<sup>2</sup> Until recently one might have objected to the portion of this research involving study of the visual system of the house cat. But recent evidence indicates choice behaviors based upon spectral discrimination can be conditioned into the cat (23).

acting channels discharging more than once increases independently of wavelength. Second, differences in CFF obtained by varying composition will be decreased as one departs from PCFs in which photic pulse durations are brief. Differences in temporal characteristics of discharge associated with wavelengths are progressively obscured by progressive increases in stimulus duration brought about by increasing PCF. The present study sought to test these suppositions.

#### D. METHOD

Two major items of equipment were used in the investigation. The first device varied the degree of spectral purity of the photic input. This apparatus was in essence a modified flicker photometer in which the two alternating components were full- and part-spectrum presentations of equal luminosity. "Maximal" purity began with presentations of the part-spectrum alone and progressed to greater and greater depurification by alternating the part-spectrum with full-spectrum in increasing proportions of the latter. The rate of alternations was made considerably greater than CFF (1280 ps), high enough to preclude effective phase interaction between the output of this mixing and the alternations produced by a second revolving disc nearer the eye.

The second item was an episcotister placed between output from the flicker photometer and the eye. This produced alternations of photic input and no input. The PCFs of .125, .250, .500, .750, and .875 were employed. The rate of this alternation was manipulated by the observer. A Weston Tachometer measured rps and hence CFF.

#### E. PROCEDURE

Two of the authors served as *Os*. Using the method of limits, CFFs of 37 separate spectral targets were determined by averaging 10 readings for each target. One target was full-spectrum, six were part-spectrum, and 30 were combinations of one of the part-spectrum luminances and the full-spectrum luminances in the following ratios: .125, .250, .500, .750, or .875. All luminosities were equated to the achromatic target set at an intensity of 4.8 c/ft.<sup>2</sup> Observations were made from a chin rest with the left eye occluded. Stimulation was foveal and near foveal, with visual angle having a value of 3° 30'. Special care was taken to eradicate all vestiges of flicker. This is a very important matter because flicker is less vigorous in part- than in full-spectrum targets and seems to differ substantially between different regions of the spectrum.



## F. RESULTS

Data are presented graphically in Figure 1 and numerically in Table 1. Each family of curves shows CFFs from one *O* for a particular wavelength distribution using all degrees of purity and all five PCFs.

PCF has the same effects for all targets of a given wavelength-achromatic balance. CFF is greatest for PCF 1/2 and declines toward PCFs 1/8 and 7/8.

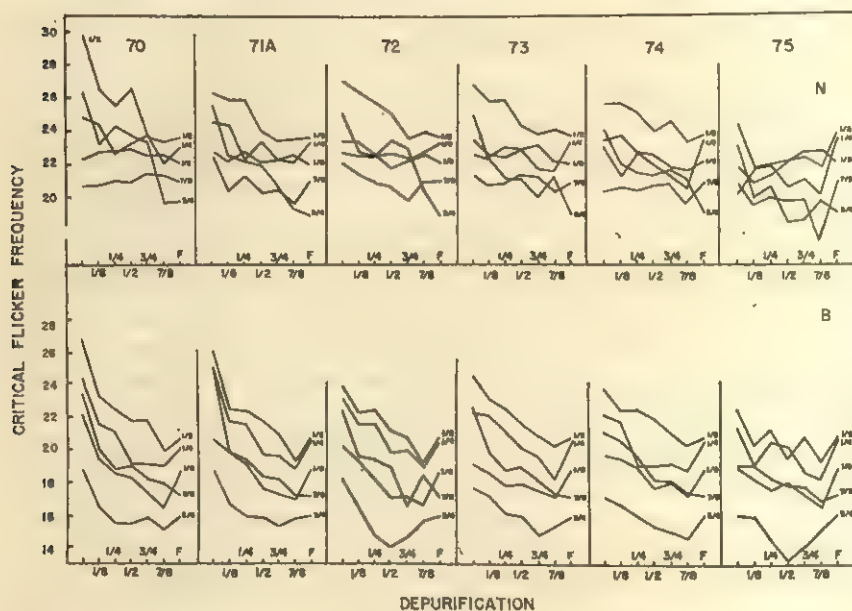


FIGURE 1

CRITICAL FLICKER FREQUENCY CURVES FOR FIVE PULSE-TO-CYCLE FRACTIONS (PCFs)  
WHEN SIX PART-SPECTRUM INPUTS ARE COMBINED WITH VARIOUS  
PROPORTIONS OF FULL-SPECTRUM INPUTS

The numbers across the top are Wratten filter numbers for the part-spectrum components. Filters 70 and 71A give red; 72 gives reddish orange; 73, yellow green; 74, green; and 75, blue green. The upper sets of curves are for one observer and the lower for the other. The fractions labelling the curves are for PCFs, the fractions along the horizontal are for various per cents of the full-spectrum components: i.e., degrees of depurification.

The rate of decline is steeper for the larger PCFs, however. Both wavelength and purity vary the range of magnitudes associated with variations in PCF. The long-wavelength end is more affected than the short, and effects decrease with decrease in purity.

The left to right course of each curve describes the effects produced by



TABLE 1  
MEAN CFF FOR SEVEN CONDITIONS OF PURITY AND SIX CONDITIONS OF WAVELENGTH

Dominant wavelength (m $\mu$ )	678	635	610	576	PCF 1/8	530	490	678	635	610	576	PCF 1/4	530	490	678	635	610	576	PCF 1/2	530	490
Mono	N	22.3	22.7	22.8	22.6	23.0	21.8	24.8	24.4	23.4	23.4	23.3	20.0	29.0	26.4	26.8	26.7	25.5	24.2		
1/8	B	22.1	20.6	20.0	19.1	19.7	19.0	24.2	25.0	23.0	22.2	22.1	21.2	26.6	25.9	24.8	24.6	23.8	22.3		
	N	22.7	22.1	22.6	22.4	21.3	21.0	24.4	24.4	23.4	22.6	23.6	21.6	26.5	25.8*	26.3	25.7	25.5	21.6		
1/4	B	19.4	19.8	19.2	18.6	19.6	18.2	21.5	21.8	22.5	22.1	21.6	19.0	23.2	22.4	22.1	23.0	22.3	20.2		
	N	22.8	21.8	22.7	23.0	22.8	21.4	22.6	22.3	22.4	22.4	22.6	21.8	25.5	25.8	25.7	25.8	25.0	21.8		
1/2	B	18.6	19.1	18.3	17.8	18.9	17.6	21.2	21.6	21.5	21.2	19.0	20.4	22.4	22.3	22.4	22.5	22.4	21.1		
	N	22.9	22.2	23.8	22.9	22.7	22.1	23.3	23.2	23.4	22.8	21.8	20.4	26.6	23.9	25.0	24.2	23.8	22.0		
3/4	B	18.3	17.6	17.1	17.9	17.6	18.0	19.1	19.6	19.8	20.1	19.0	20.5	21.7	21.7	21.1	21.6	21.9	19.4		
	N	22.5	22.3	22.3	23.0	22.0	22.8	23.8	22.2	23.0	21.6	21.2	20.8	23.6	23.4	23.4	23.7	24.4	22.2		
7/8	B	17.4	17.4	17.2	17.5	18.0	17.2	19.1	19.6	19.9	19.6	19.2	18.5	21.8	20.9	20.7	20.8	21.0	20.7		
	N	22.5	21.6	22.6	22.2	21.7	22.8	22.0	22.2	21.2	21.6	20.4	20.0	23.4	23.5	23.9	24.0	23.2	21.6		
Full	B	16.4	17.0	16.7	17.2	17.1	16.4	19.0	18.8	19.2	18.2	18.6	18.1	21.8	20.9	20.7	20.8	21.0	20.7		
	N	22.0	22.0	22.0	22.0	22.0	22.0	23.3	23.3	23.3	23.3	23.3	23.3	23.6	23.6	23.6	23.6	23.6	23.6		
	B	18.7	18.7	18.7	18.7	18.7	18.7	20.5	20.5	20.5	20.5	20.5	20.5	20.7	20.7	20.7	20.7	20.7	20.7		

TABLE 1 (continued)

Dominant wavelength (m $\mu$ )	678	635	610	576	PCF 3/4	530	490	678	635	610	576	PCF 7/8	530	490
Mono	N	27.2	26.6	26.0	26.0	25.0	24.0	20.7	22.4	22.1	21.4	20.5	20.8	
1/8	B	23.6	24.9	22.4	22.5	21.0	19.0	18.7	18.7	18.2	17.6	17.1	16.0	
	N	24.2	23.5	23.7	22.4	23.0	20.8	20.7	20.4	21.4	20.8	20.6	19.6	
1/4	B	20.0	19.8	19.6	19.8	20.6	19.0	16.5	16.6	16.6	17.2	16.6	16.0	
	N	25.3	23.2	23.5	22.2	22.4	21.4	21.1	21.3	20.8	20.8	20.5	20.0	
1/2	B	18.3	19.4	19.4	18.8	19.6	18.1	15.6	16.0	14.8	16.2	16.0	14.4	
	N	24.8	23.0	22.8	22.0	22.2	19.4	20.9	20.2	20.9	21.4	20.8	19.8	
3/4	B	19.0	18.4	19.0	19.0	19.2	17.8	15.6	15.9	14.2	15.6	15.4	13.2	
	N	24.4	21.8	23.2	21.0	22.5	19.5	21.6	21.4	19.8	21.3	21.0	19.8	
7/8	B	18.2	18.2	16.6	18.2	18.2	17.9	15.9	15.4	14.8	14.9	15.0	14.2	
	N	20.6	20.4	22.2	22.2	22.0	20.6	21.4	19.6	21.0	20.4	19.7	17.4	
Full	B	17.9	17.2	18.5	17.6	17.4	16.8	15.2	16.4	15.5	13.6	14.8	15.2	
	N	19.9	19.9	19.9	19.9	19.9	19.9	21.1	21.1	21.1	21.1	21.1	21.1	
	B	17.2	17.2	17.2	17.2	17.2	17.2	16.0	16.0	16.0	16.0	16.0	16.0	

reduction of purity. The range of differences are often small, but the directions of the differences are consistent. In general, CFF decreases as purity decreases. When the CFF for full-spectrum is compared with that for the most restricted wavelength of each series, the full- exceeds the part-spectrum in less than 10 per cent of the cases. This drop is clearly apparent when data for each filter at each PCF is rank ordered. Rank totals show a decrease in CFF with decrease in purity. But not all parts of the spectrum nor PCFs are equally affected. The shortest wavelength manifested the smallest range of CFFs. A monochromatic to full-spectrum decrease occurred only half as frequently for Wratten 75 (blue) as for Wratten 70 (red). Large and small PCFs have the smallest range of full-spectrum-part-spectrum differences. Larger PCFs have a greater tendency to disturb the general order of decrease in CFF than the smaller, however.

The data suggest an effect of wavelength upon CFF. When narrowest parts of the spectrum experimented upon are compared, CFFs decrease from red to blue. This is true in nine of 10 cases. The range of differences are of moderate magnitude: not in any case over six cps. The greatest difference in red-blue CFF occurs with PCF 1/2. Differences at the extremes (PCF 1/8 and 7/8) do not exceed two cps and are nonexistent for one *O*. As might be expected, achromatic mixing reduces the likelihood of a red-blue order. Nine of 10 of the unmixed curves show a red-blue decrease, but only six of 10 do so when the input is 7/8 achromatic.

### G. DISCUSSION

The results show that CFF is affected by variations of PCF, purity, and wavelength. These effects are predictable, once it is presumed that the cortical pattern associated with each specific type of sensory end result is temporally unique.

As suggested before, the fact that wavelength varies CFF seems to be related to the fact that various wavelengths involve different latencies in arousing sensory response. The general reduction in CFF with reduction in purity suggests that when portions of the pathway are mediating different hues, the neural discharge and recovery characteristics become so heterogeneous as to increase the uniformity of activity throughout each cycle of repetition. Thus with depurification a smaller increase in rate is required to make the density of activity uniform enough to yield a constant level of brightness (fusion).

It cannot be presumed, however, that simply because discharge density was always involved, all of the CFF phenomena described were due to a single

set of factors. Total flux per cycle was also varied by PCF. The decline in CFFs above PCF  $1/2$  may perhaps be due to the relatively short recovery period between inputs; and the drop in CFF below  $1/2$ , due to the reduction in total flux per cycle. These possibilities should be directly investigated.

#### H. SUMMARY

Two Os viewed 37 targets varying in wavelength and purity and PCF through an episcotister. Target fields were equated in luminosity using the method of flicker. CFFs were obtained for each color condition using five pulse-to-cycle fractions (PCFs). Data showed CFF to be varied by PCF, purity, and wavelength. Temporal theory can be employed in explaining the results.

#### REFERENCES

1. ALLEN, F. Variation of visual sensory reflex action with intensity of stimulation. *J. Opt. Soc. Amer.*, 1926, **13**, 383-430.
2. BALL, R. J. An investigation of chromatic brightness enhancement tendencies. *Amer. J. Optom.*, 1964, **41**, 333-361.
3. BARTLEY, S. H., & NELSON, T. M. Certain chromatic and brightness changes associated with rate of intermittency of photic stimulation. *J. of Psychol.*, 1960, **50**, 323-332.
4. CHANG, H. T. Functional organization of visual pathways. *Res. Pub. Assoc. Nerv. & Ment. Dis.*, 1950, **30**, 430-433.
5. COHEN, J., & GORDON, D. The Prevost-Fechner-Benham subjective colors. *Psychol. Bull.*, 1949, **46**, 97-136.
6. CROZIER, W. J., & WOLF, E. Theory and measurement of visual mechanisms: VI. Wavelength and flash duration in flicker. *J. Gen. Physiol.*, 1941, **25**, 89-110.
7. FERREE, C. E., & RAND, G. A preliminary study of deficiencies of the method of flicker for the photometry of lights of different colors. *Psychol. Rev.*, 1915, **22**, 110-162.
8. FRY, G. A. A photo-receptor mechanism for the modulation theory of color vision. *J. Opt. Soc. Amer.*, 1945, **35**, 114-135.
9. HECHT, S., & SHLAER, S. Intermittent stimulation by light: V. The relation between intensity and critical frequency for different frequency for different parts of the spectrum. *J. Gen. Physiol.*, 1936, **19**, 965-979.
10. IVES, H. E. Studies in the photometry of lights of different colours. *Philos. Mag. & J. Sci.*, 1912, **24**, 149-188.
11. ———. Studies in the photometry of lights of different colours: II. Spectral luminosity curves by the method of critical frequency. *Philos. Mag. & J. Sci.*, 1912, **24**, 352-370.
12. IVES, H. E., & KINGSBURY, E. F. The theory of the flicker photometer. *Philos. Mag. & J. Sci.*, 1914, **28**, 708-728.
13. ———. Theory of the flicker photometer: II. Unsymmetrical conditions. *Philos. Mag. & J. Sci.*, 1916, **31**, 290-321.
14. LANDIS, C. Determinants of the critical flicker-frequency threshold. *Physiol. Rev.*, 1954, **34**, 259-286.
15. LENNOX, M. A. Geniculate and cortical responses to colored light in the cat. *J. Neurophysiol.*, 1956, **19**, 271-279.

16. ———. The on responses to colored flash in single optic tract fibers: Correlation with conducting velocity. *J. Neurophysiol.*, 1958, **21**, 70-84.
17. MADSEN, A., & LENNOX, M. A. Cortical and retinal responses to colored light flash in the anesthetized cat. *J. Neurophysiol.*, 1955, **18**, 412-424.
18. NELSON, T. M., & BARTLEY, S. H. The role of PCF in temporal manipulations of color. *J. of Psychol.*, 1961, **52**, 457-477.
19. NELSON, T. M., BARTLEY, S. H., & BLECK, F. C. CFF and spectral purity. (To be published.)
20. NELSON, T. M., BARTLEY, S. H., & FORD, Z. The brightness of part-spectrum targets stimulating the eye parafoveally at intermittency rates below fusion. *J. of Psychol.*, 1963, **55**, 387-396.
21. NELSON, T. M., BARTLEY, S. H., & MACKAVEY, W. R. Responses to certain Pseudo-isochromatic charts viewed in intermittent illuminance. *Percept. & Motor Skills*, 1961, **13**, 227-231.
22. PIERON, H. Loi de la vitesse d'établissement des processus chromatiques fondamentaux en fonction de l'intensité de l'excitation lumineuse. *C. R. Soc. Biol.*, 1922, **174**, 1294-1296.
23. SECHZER, J. A., & BROWN, J. S. Color discrimination in the cat. *Science*, 1964, **144**, 427-429.
24. STEWART, G. N. Is the law of Talbot true for very rapidly intermittent light? *Proc. Roy. Soc. Edinb.*, 1887-1888, **15**, 441-464.
25. TROLAND, L. T. Notes on flicker photometry: Flicker-photometer frequency as a function of the color of the standard and measured light. *J. Franklin Inst.*, 1916, **181**, 853-855.
26. ———. The enigma of color vision. *Amer. J. Physiol. & Opt.*, 1921, **2**, 23-48.
27. ———. The Principles of Psychophysiology: Vol. II. Sensation., New York: Van Nostrand, 1930.
28. TRUSS, C. V. Chromatic flicker fusion frequency as a function of chromaticity. *J. Opt. Soc. Amer.*, 1957, **47**, 1130-1134.

*Department of Psychology*  
*Michigan State University*  
*East Lansing, Michigan*

## SUBJECTIVE COLOR: A NEW METHOD OF PRODUCING THE PHENOMENON\*

*Department of Psychology, Hunter College of the City  
University of New York*

---

CHARLES A. KNEHR AND ROBERT J. LORENZ

---

### A. INTRODUCTION

The term "subjective color," in a sense, is a misnomer since any color response depends upon the functional characteristics of the organism in respect to a portion of the radiant-energy spectrum. Historically, the term has been applied to the resulting perception of color when patterns of black-and-white stimuli are in motion (usually being rotated) in white light, a phenomenon that was first discovered well over one hundred years ago. Cohen and Gordon (1) have provided an excellent historical review of the phenomenon.

While attempting to account for previously reported individual differences in subjective color in terms of certain hypothetical visual-frequency functions (2), one of the authors (R.J.L.) tried illuminating the rotating pattern with the white-light pulses produced by a flicker-fusion threshold apparatus. The resulting bright and saturated colors were clearly a function of the combination of disc rotation and light-pulse frequencies. Further work produced the following observations.

### B. PROCEDURE

Since the flicker apparatus could not be set to pulse steadily at particular frequencies, some other means of controlling light frequencies was required. A white-light source (Sylvania 1131C), which was pulsed by a Strobotac,<sup>1</sup> failed to give the color effect while producing the usual stroboscopic effects. With an apparatus providing longer and variable "on" time during light pulses of square-wave character, again no color effect was apparent.

Then it was noted that, by manually flickering the fluorescent lamps operating from a direct-current power supply built into the subjective-color apparatus, the color effect was again apparent. The light source controlled by the flicker apparatus is also a small fluorescent lamp. A standard six-watt fluorescent lamp was then suitably connected to the output of the Strobotac for further study of the effect. By mechanically pulsing the light from a

---

\* Received in the Editorial Office on June 9, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Product of General Radio Corporation.



tungsten projection lamp so as to approximate roughly the rise and decay characteristics of the fluorescent lamp, the same color effects were obtained.

The essential situation seems to be an advancing front moving across successive retinal elements during the exponential decay phase of fluorescent illumination following the sharp rise of a short light pulse (about 0.6 msec). The simplest case is provided by rotating a half-white-and-half-black disc. White following black in the direction of rotation is, in effect, an advancing "on" situation across the receptors during the brightest part of the decay interval. Black following white is an advancing "off" situation. For both CAK and R.J.L., the advancing "on" produces an orange color overlying the black where that was positioned during the light pulse. Ahead of the advancing "off" (black) front a well-saturated blue is seen. Both colors gradually become unsaturated and then disappear in the direction of rotation. When the disc speed and the light frequency are the same, these patterns remain stationary, thereby contributing to easy observation.

With more complicated disc patterns, such as the one shown in Figure 1, the following results have been obtained consistently. When the rotation and

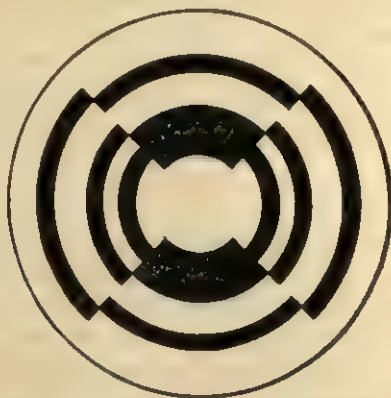


FIGURE 1  
A COMPLICATED DISC PATTERN

light-pulse frequencies are the same, narrow saturated sectors of these colors are seen at a rate of one or two cycles per second. Increasing the frequency of rotation and corresponding light-pulse frequency merely increases the angular size of the sector in which the color is seen with a concomitant decrease in saturation. Pulsing the light source at frequencies above the speed of disc rotation yields more complex patterns involving these same colors, with the

following exception. Varying with the disc pattern, when disc and light-pulse frequencies differ in certain ratios, the perceived colors disappear and are replaced by a clean, colorless gray. This loss of color apparently occurs when the two observed color processes are phased in and out alternately at a rapid rate, thus superimposing these two effects on the same retinal areas.

Discs in which these "on" and "off" effects occur adjacent to and on opposing sides of narrow  $360^\circ$  bands of white and black seem to show no induction or simultaneous contrast effects. A white disc prepared with  $20^\circ$  bands of blue, green, orange, yellow, and red yielded the same colors as before, with blue extending into the white ahead of the bands and the orange color spreading over the painted bands, which were of lower brightness than the white disc. This phenomenon appears to result from advancing changes in brightness levels and is not significantly altered by varying the color of the illumination. It has been noted before that subjective colors are not appreciably altered by using monochromatic light.

### C. DISCUSSION

By the expedient of producing a variety of discs with regular and irregular patterns illuminated with trains of single pulses at different frequencies, it has not been possible to duplicate the range of colors obtained under varying conditions with a Benham disc in continuous white light. By using pairs of pulses, the number of colors perceived can be increased considerably. Adequate exploration of this finding will require sophisticated equipment for the control of the several variables involved.

It was also noted that when the lamp being pulsed was moved rapidly, the colors seen were blue and yellow irrespective of the frequency of light pulses or the speed of movement. The colors seen were subjectively matched using a monochromator. The readings for the two observers were 590 and 601  $m\mu$  for the yellow and 481 and 491  $m\mu$  for the blue.<sup>2</sup> These are close to the matching ranges for the yellow and blue psychological primaries reported by other workers (3). A comparison of the blue color obtained with the moving pulse light source and the blue seen on the disc suggests that the observer is experiencing the same color for both conditions. The orange color seen on the disc might be some function of the yellow superimposed on the black.

To our knowledge, colors seen on patterned black-and-white discs rotating in pulsed white light have not been reported before. What we may have dis-

---

<sup>2</sup> Mr. Paul Balkan of the Hunter College Department of Physics very kindly provided the instruments for these measurements in addition to measuring the characteristics of the light source.

covered, so far, is a technique for breaking down complex subjective color events into repeated short-time samples for at least some of the processes involved. The colors can be seen in relation to pattern characteristics without the usual interference of mottling and other effects especially noticeable at slow speeds on such discs spun in steady illumination. Conceivably, this technique could lead to a new approach to a long refractory problem. It seems possible that these data may have some relevance to the more general theory of color vision.

#### REFERENCES

1. COHEN, J., & GORDON, D. A. The Prevost-Fechner-Benham subjective colors. *Psychol. Bull.*, 1949, **46**, 97-136.
2. KNEHR, C. A. Individual differences in subjective color. *J. of Psychol.*, 1953, **36**, 289-294.
3. PICKFORD, R. W. Individual Differences in Color Vision. New York: Macmillan, 1951, P. 15.

*Hunter College*  
*Bedford Park Boulevard West*  
*Bronx 68, New York*

## TRAIT JUDGMENTS OF MILITARY PILOTS FROM PHOTOGRAPHS\*

*Department of Psychology, Boston College*

---

D. J. BAER

---

### A. PURPOSE

Since the word "stereotype" was introduced by Lippman (4) to represent rigid and standardized "pictures in our head," many studies (1, 2, 3, 5, 7) have illustrated how people tend to attribute traits to individuals on the basis of the traits that are assumed to characterize the group to which the individuals belong.

Among the more "glamorous" professions, perhaps the Air Force pilot has received much favorable publicity from the mass media of communication; yet relatively few people have had direct personal contact with such an individual. This lack of contact would seem to favor the emergence of a stereotyped image of what a "good pilot" should look like.

The purpose of the present study was to examine the consistency in the ranking of behavioral traits of pilots when such ranking is based on appearance in photographs and to relate the ranking to the physical characteristics of the pilot. On the basis of the type of military pilots portrayed in recruitment posters, motion-picture films, and the finding of Secord (6) that people attribute some of the traits of familiar persons to strangers whose photographs resemble them (whether or not they are aware of the connection), it was hypothesized that younger, taller, and thinner pilots would be ranked as "better pilots," and that older, taller, and heavier pilots would be ranked as "better leaders."

### B. Method

#### 1. Subjects

Thirty male college students who had never met an Air Force pilot and who had little knowledge of flying skills or military life (other than the information acquired from their exposure to the mass media of communication) were selected as Ss.

---

\*Received in the Editorial Office on June 12, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

## 2. Test Instrument

The photographs of 10 military pilots, each wearing a uniform flight suit with no military rank displayed and posed in a standard position beside a T28 training plane, were used as test stimuli. The pilots ranged from 24 to 42 years in age, from 160 to 195 pounds in weight, from 66 to 72 inches in height, and from lieutenant to lieutenant colonel in military rank.

## 3. Procedure

The photographs were simultaneously presented to each *S* with the instructions to rank the pilots according to what he considered the "best pilot." No reference was made as to what a "best pilot" should be; each *S* was required to use his own frame of reference. This same procedure was repeated two more times, using the categories of "best leader" and "most aggressive." The 10 photographs were identified from A to J and were in a different randomized order each time when presented to the *Ss*.

## C. RESULTS AND DISCUSSION

The consistency of the rankings by the 30 *Ss* for the three behavior traits was determined by Kendall's coefficient of concordance, *W*. It may be seen in Table 1 that the chi-square tests were significant beyond the .01 level for each behavior trait. Such high consistencies seem to justify the assignment of a rank based on the composite ranks assigned (by the 30 *Ss*) to each photograph for each behavior trait.

TABLE 1  
COEFFICIENTS OF CONCORDANCE, *W*, AND CHI SQUARES BASED ON TESTS OF SIGNIFICANCE  
OF CONSISTENCY OF RANKING OF 10 PILOT PHOTOGRAPHS FOR  
THREE BEHAVIOR TRAITS BY 30 *Ss*

Trait	<i>W</i>	Chi square
Best pilot	.22	41.02*
Best leader	.45	81.00*
Most aggressive	.20	35.81*

\*  $p < .01$ ,  $df = 9$ .

In Table 2 are presented the Spearman rank-difference correlations between the rank order of the physical and behavior traits associated with the pilots. The significant positive relationships found between the physical and behavioral rankings tend to support the hypotheses that the heavier pilots are considered the "better pilots," and that the older pilots are considered the "better leaders."

The finding that the older pilots tended to be rated as the better pilots was



opposite to the result that had been predicted. Perhaps the term "best pilot" connoted to the Ss someone who has had longer experience in his career rather than someone who may possess a greater psychomotor proficiency because of his youth. The negative rho of  $-.12$  between height and aggressiveness, although insignificant, seems consistent in direction with the common

TABLE 2  
SPEARMAN RANK-DIFFERENCE CORRELATIONS BETWEEN RANKED PHYSICAL TRAITS AND  
ASCRIBED BEHAVIOR TRAITS OF PILOTS BASED ON PILOT PHOTOGRAPHS

Trait	Best pilot	Best leader	Most aggressive
Age	.54	.75**	.02
Height	.42	.16	$-.12$
Weight	.61**	.26	.59*
Best pilot		.84**	.03
Best leader			$-.06$

\*  $p < .05$  (one-tailed test).

\*\*  $p < .01$  (one-tailed test).

stereotype that shorter men are considered more aggressive. Perhaps the insignificant correlation between "best pilot" and height was due in part to the short range of heights of the pilots. Even so, the correlation was in the expected direction and of a large magnitude, reaching the .06 level of significance. The fact that not all of the correlations between the behavior traits were significant indicates that a consistent halo effect did not influence the Ss' judgments of the photographs for all of the categories.

#### D. SUMMARY

Ten photographs of military pilots were ranked by 30 Ss for three behavior traits: "best pilot," "best leader," and "most aggressive." The significant consistencies found in the rankings of the Ss for each behavior trait show that a stereotyped image existed for the three behavior traits. Significant correlations were found between the ranked physical traits and the ascribed behavioral traits of the pilots—with the older, taller, and heavier pilots being ranked as the "better pilots" and as the "better leaders."

#### REFERENCES

1. BRUNER, J. S., & PERLMUTTER, H. V. Compatriot and foreigner: a study of impression formation in three countries. *J. Abn. & Soc. Psychol.*, 1957, **55**, 253-260.
2. CAMPBELL, D. T., & BURWEN, L. S. Trait judgments from photographs as a projective device. *J. Clin. Psychol.*, 1956, **12**, 215-221.
3. EDWARDS, A. E. Four dimensions in political stereotype. *J. Abn. & Soc. Psychol.*, 1940, **35**, 566-572.
4. LIPPMAN, W. *Public Opinion*. New York: Harcourt, Brace, 1922.

5. RAZRAN, G. Ethnic dislikes and stereotypes: A laboratory study. *J. Abn. & Soc. Psychol.*, 1950, **45**, 7-27.
6. SECORD, P. F. Facial features and inference processes in interpersonal perception. In Tagiuri, R., & Petrullo, L. (Eds.), *Person Perception and Interpersonal Behavior*. Stanford, Calif.: Stanford Univ. Press, 1958. Pp. 300-315.
7. THIBAUT, J. W., & RIECKEN, H. W. Some determinants and consequences of the perception of social causality. *J. Personal*, 1955, **24**, 113-133.

*Department of Psychology*  
*Boston College*  
*Chestnut Hill 67, Massachusetts*

## CONFLICT MODELS, GAMES, AND DRINKING PATTERNS\*

*Veterans Administration Hospital, Brockton, Massachusetts*

---

HENRY S. G. CUTTER

---

### A. INTRODUCTION

This paper derives a small group-level model from individual-level conflict models on the one hand and from a cross-cultural-level theory on the other. It is hoped that the model of the game and its players will aid both the analysis of social-drinking behavior and the synthesis of theories of the individual with theories of society.

The discussion of controlled- and uncontrolled-drinking behavior employs Heilizer's conflict-of-motives models (2) and Caillois' theory of play and games (1).

It is proposed that controlled drinking shares many of the characteristics of games as defined by Caillois (1) and that out-of-control drinking lacks many of these game-like characteristics.

Heilizer's concept of intrinsic motive (organism-based stimuli that impel responses from within the individual) and his concept of extrinsic motive (environment-based stimuli that impel responses from outside the individual) are used to explicate Caillois' game concepts. They are then extended to stimulus situations that involve no conflict about drinking alcoholic beverages.

Some practical implications for the prevention of uncontrolled drinking are mentioned.

### B. DRINKING BEHAVIOR AND CONFLICT

Alcoholism rates are higher among populations that have negative attitudes toward alcohol consumption and whose members proportionately drink little as compared to those in populations whose members have favorable attitudes toward alcohol and drink ubiquitously (4). There are some exceptions that must be considered and explained by the drinking models presented below.

Heilizer (2) has developed two formal models that aid in understanding the relevant variables in controlled and uncontrolled drinking. He posits that two opposing motives are important in analyzing drinking behavior. One leads to the positive approach to alcohol; the other leads to the negative

---

\* Received in the Editorial Office on June 17, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

avoidance of alcohol. These two motives can operate both within the individual as an intrinsic motive or attitude (either positive or negative) acquired during periods of earlier socialization and as an extrinsic motive or social pressure (either positive or negative) that can operate on the person in his contemporary relations to a face-to-face drinking group.

Intrinsic motives produce gradually strengthening responses to a goal. The stimuli produced by intrinsic motives operate from within and thus are effective even when the individual is at a considerable distance from a goal.

Extrinsic motives produce sharply strengthening responses to a goal. The stimuli produced by extrinsic motives are based in the environment and can only operate on the individual when his receptors are within range of the stimuli and the individual is therefore relatively close to a goal.

Heilizer's controlled-drinking model pairs an intrinsic positive motive to drink with an extrinsic negative motive to avoid drinking. This implies that the person thus motivated will seek a drink even when it is not available in the immediate vicinity. When such a person has found his drink he will drink in a gradual, steady fashion until extrinsic negative motives (such as group social censure for overindulgence) interfere with his continuing drinking.

Heilizer's out-of-control-drinking model pairs an intrinsic negative motive to avoid drinking with an extrinsic positive motive to drink. This implies that the person motivated by this pattern will attempt to avoid drinking or drinking situations even when they are remote. This individual has to be forced to drink by social pressure. When he does drink, his drinking is rapid and compulsive and is terminated only by inebriation. The alcohol reduces his tension of having been forced to drink and reinforces the habit of drinking when in conflict about drinking.

#### C. DRINKING BEHAVIOR WITH NO CONFLICT

There is no logical or theoretical reason why a positive intrinsic motive cannot be paired with a positive extrinsic motive or a negative intrinsic motive cannot be paired with a negative extrinsic motive. The latter, unconflicted combination of negative motives is interesting because it describes teetotalers who, if they drink, have a high probability of becoming alcoholics. The former combination of positive unconflicted motives is interesting because it describes the French pattern of drinking where drinking is ubiquitous and there is a high rate of alcoholism (4).

#### D. MOTIVE PATTERNS AND GAMES

It would appear, then, that uncontrolled drinking is not always a function of the conflict of motives. Cailliois (1), in his analysis of play and games,

states that the function of the drinking or "vertigo" game is to permit the experience, the expression, and the release of normally inhibited behaviors and affects under conditions that protect the players. Drinking as a game may also facilitate gregariousness and sociability, as well as permit people to practice certain social skills within the small face-to-face drinking group as a by-product. When the game of drinking fails to protect the players from the damaging consequences of drinking, then it ceases to be play. An analysis of the characteristics of games may help in understanding how games perform these rather contradictory functions.

Caillois (1, pp. 9-10) characterizes play and games as:

1. *Free*: in which playing is not obligatory, if it were it would at once lose its attractive and joyous quality as a diversion.

By free, Caillois implies that play does not depend on positive extrinsic motives or rewards. Rather the player is intrinsically positively motivated and rewarded by the play activity itself. Intrinsic negative motivation is inferentially excluded.

2. *Separate*: circumscribed within limits of space and time, defined and fixed in advance.

Separateness is an extrinsic negative motive in the form of limits or controls.

Depending on the society involved, drinking may be spatially restricted to bars or parties, or temporally restricted to certain times of day, the weekends, or annual carnivals.

3. *Uncertain*: the course of which cannot be determined nor the results obtained beforehand, and some latitude for innovation being left to the player's initiative.

Uncertainty of outcome is more relevant to gambling and strategy games than to drinking. Variability of response is important in all forms of play. If games are excessively ritualized they cease to be recreation and become religious or sacred ceremonies.

4. *Unproductive*: creating neither goods nor wealth, nor new elements of any kind; and except for the exchange of property among the players ending in a situation identical to that prevailing at the beginning of the game.

This definition states that the reward of extrinsically motivated responses is ungame-like. Amateur as opposed to professional participation in sports is differentiated on this criterion.

5. *Governed by rules*: under conventions that suspend ordinary laws, and for the moment establish new legislation which alone counts.



Caillois implies that rules have both an intrinsic positive motive quality and an extrinsic negative motive quality. Rules are intrinsically positive in the sense that if one is intrinsically motivated to play the game and breaks the rules, then one is failing to play the game. Usually, rules act as limits in an extrinsic negative motive sense.

For example, the suburban-cocktail-party rules encourage flirting with others' mates. The women are enjoined from comparing detergents and babies, and the men are chastized if they lapse into "shop" talk. The game functions to make the wives feel that they are still attractive to other men, while husbands obtain a brief respite from monogamy without suffering the consequences of marital infidelity.

6. *Make believe*: accompanied by a special awareness of a second reality or of a free unreality, as against real life.

If the flirtation at the hypothetical cocktail party is not seen as "make-believe," a residue of jealousy, marital strife, and emotional disturbance is left. The severe breakdown of the "make-believe" attitude differentiates psychotics (who confuse their fantasies with external events) from painters (who carefully label their imaginative productions "art") or from normals (who label their fantasies daydreams).

In game terms, controlled drinkers are amateurs. They drink freely or voluntarily. They drink in places separate from the real world in bars, pubs, and at parties and at specified times that do not conflict with or become confused with the "other world." This drinking is unproductive and is carried on for its own sake. It is conducted in such a fashion that negative consequences are also avoided. Drinking is conducted according to rules and conventions that tend to control the drinking activity.

Conversely, out-of-control drinkers are professionals. Their drinking is confused with and affects the "reality" of work and family, both in terms of place and time. Liquor is concealed in lockers and is consumed during working hours. Drinking is seen as productive by the professional drinker, in that it is consumed for some reward. Drinking ceases to be unreal and becomes the "for-real," principal goal of the individual and is confused with and conflicts with the more usual realities of work and family. Drinking is not conducted according to the clear rules and rituals of games, but is conducted according to the confused and contradictory rules of life or "reality."

If one places the various combinations of intrinsic and extrinsic positive and negative motive pairings in the order of least to most game-like, the following ranking of pairings is obtained:

Intrinsic negative—Extrinsic positive  
Intrinsic negative—Extrinsic negative  
Intrinsic positive—Extrinsic positive  
Intrinsic positive—Extrinsic negative

The top motive pairing is identical to the out-of-control conflict model and is the least game-like because it violates two of the criteria characterizing a game, while each of the two middle pairings violates one criterion and the bottom pairing is identical to the controlled-drinking conflict model and is the most game-like because it fits most of the criteria for games.

From the above discussion, it seems reasonable to argue that drinking is a controlled activity when it closely fits the criteria of a game and appears to be less controlled when it violates the game criteria.

#### E. DRINKING BEHAVIOR, GAMES, AND THE PLAYERS

It is now possible to trace the operation of these motive patterns, to extend the terminology of play and games to include the variables just discussed, and to trace the conditions that lead to uncontrolled drinking. The intrinsically positively motivated drinker, or amateur drinker, drinks within the rules of the drinking game as they pertain to the time, place, and mode of drinking. If the nondrinking spoilsport, the subject with an intrinsic negative motive toward drinking coming from an extrinsic negative environment, wants to join in the drinking game and share in the sociability and affect expression accompanying it, he is in a dilemma. The positive intrinsically motivated behavior of the amateur drinker functions as an extrinsic positive motive to drink countering the spoilsport's intrinsic negative motive not to drink. If he drinks, he has resolved the interpersonal conflict implied in trying to be a member of a group the central activity of which he specifically rejects, but he has opened up an intrapersonal conflict in that he has engaged in behavior that conflicts with his intrinsic negative or antidrinking motive. The amateurs may respond to the spoilsport's nondrinking behavior as a criticism of their game, leading them to apply extra extrinsic positive motive pressure to overcome his intrinsic negative motives against drinking. If the spoilsport drinks, he becomes a professional drinker. Professional drinkers have trouble staying in the amateur drinking group because they tend to become out-of-control drinkers because of their lack of extrinsic negative motives or rules.

The professional drinker, dismissed from the amateurs, may attain a stable role by joining other professionals. The professional who finds life among other professionals unpalatable may want to rejoin his original tectotal group.

Violation of their most basic tenets may make him unacceptable to them. Social acceptance may be gained at the price of self-esteem by becoming a professional confessor, punishing himself, and being punished for violating his own and the group's norms. Attempts to conceal his earlier drinking always leaves the fear of being suddenly discovered and again becoming alienated from a group.

Alcoholics Anonymous encourages self-critical confession in order to strengthen one's own intrinsic negative motives and others' extrinsic negative motives. This implies that while some A. A. members will never drink again (since they will avoid alcohol), the other A. A. members who do drink are likely to drink more than they did previously in order to reduce the self-criticism-strengthened intrinsic negative drinking motives.

The other type of professional drinker is both intrinsically and extrinsically positively motivated to drink. In this category, bartenders and the French drinker (4) have a combination of motives that permits no separation of work and play. If this type of drinker wants to stop drinking recreationally to avoid excess, he is likely to be subject to the same pressures as the nondrinking spoilsport.

Drinking for many Frenchmen is a constant accompaniment of other activities in the same way that cigarets are in American culture. This continuous morning-to-night drinking produces few of the overt symptoms of intoxication observed in the American alcoholic (4). The French alcoholic is usually diagnosed by physiological criteria rather than by psychological or social criteria. Until the onset of brain damage, the French alcoholic may appear to be the perfect family man and breadwinner. It is the presence of the extrinsic positive motive to drink, rather than conflict, that induces a lack of separation of the drinking situation from the rest of life and that leads to this form of alcoholism.

#### F. PRACTICAL IMPLICATIONS OF THE MODELS

The models presented have some practical implications for controlling drinking behavior.

First, the use of scare communications by temperance organizations are not only useless but dangerous since they reinforce intrinsic negative motives toward alcohol, thus increasing the level of conflict when faced with the dilemma of drinking or not drinking.

Second, the models suggest that adolescent drinking should be encouraged at appropriate times and places. As in sports, a coach could guide the neophyte drinkers in mastering the amateur drinking game. This principle has been

applied in California to another exciting but potentially dangerous activity by providing hot rodders with legitimate drag strips and by encouraging the formation of strictly self-regulating clubs.

Third, the models suggest that all advertisements of alcoholic beverages should be abolished to reduce the extrinsic positive motive pressure to drink. Some advertisements are particularly pathogenic, according to Hoffman (3), because they equate drinking with social acceptability, happiness, and work effectiveness.

Fourth, it is implicit in the models discussed that intrinsically negatively motivated responses to alcohol are usually overcome by external social pressure. The teetotaler drinks only when the group forces him to. It follows that people who succumb readily to group pressure to drink are more likely to become alcoholics than those who do not succumb readily, holding constant the strength of the aversive responses to alcohol.

Fifth, the question whether a professional drinker, unlike the professional athlete, can become an amateur is worth exploring. If it is possible to bring uncontrolled drinking under control, the conflict model and game-theory criteria should provide procedural guidelines.

#### G. SUMMARY

Controlled- and uncontrolled-drinking behavior was discussed as a function of the absence or presence of two types of motive conflict which were then related to a theory of games. The games analysis of drinking behavior was extended to certain unconflicted drinking patterns. Some practical implications for the control of drinking behavior were discussed.

#### REFERENCES

1. CAILLOIS, R. *Man, Play, and Games*. Glencoe, Ill.: The Free Press, 1961.
2. HELLIZER, F. Conflict models, alcohol, and drinking patterns. *J. of Psychol.*, 1964, 57, 457-473.
3. HOFFMAN, J. Mass media and drinking patterns. Unpublished manuscript, Veterans Administration Hospital, Brockton, Massachusetts, 1962.
4. JELLINEK, E. M. *The Disease Concept of Alcoholism*. New Haven: Hillhouse Press, 1960.

*Veterans Administration Hospital  
Brockton 5, Massachusetts*





## ITEM CONTEXT AND POSITION EFFECTS ON MULTIPLE-CHOICE TESTS\*

*Department of Psychology, University of North Carolina at Greensboro*

---

LEWIS R. AIKEN, JR.

---

### A. INTRODUCTION AND PURPOSE

Since Cronbach's original paper on response sets (2), the conditions of occurrence of such sets and their effects on responses to various kinds of questionnaire and test items have been studied extensively. Cronbach has concluded (3), and it appears to be generally acknowledged, that forced-choice and multiple-choice items are less influenced by response sets than are other types of items. In regard to the "position-response set," in particular, a number of investigations indicate that neither the serial position of a multiple-choice item in a test (5) nor the position of the correct alternative within the item (4, 6) has a significant effect on the response to the item.

It has been demonstrated, however, that certain contextual factors may affect responses to multiple-choice items. For example, the results of a study by Carr and Sax (1) showed that more items were attempted and higher scores were obtained on the spiral-omnibus form than on the subtest form of the intelligence test which they administered. These findings indicate that the tendency to respond (or not to respond) to specified multiple-choice items may be influenced not only by the difficulty of the items but also by the difficulty context in which the items are embedded. Therefore, if examinees are permitted unlimited time in which to answer all items on a multiple-choice test, one measure of the effects of item-difficulty context on responses would be the number of "critical" items, or very difficult items common to all tests, left blank in tests of varying difficulty levels. This was the criterion used in the present investigation for studying the effects of difficulty context on responses to multiple-choice test items. Of secondary interest were the effects of varying the position of the selected "critical" items on whether or not these items are attempted.

### B. METHOD

A limited investigation of the effects of item-difficulty context and item position on responses to nine "critical," four-alternative, multiple-choice

---

\* Received in the Editorial Office on June 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

vocabulary items was conducted. A pool of 135 "noncritical" items was taken from the Elementary, Junior High School, Secondary, and Advanced 1957 short forms of the California Test of Mental Maturity (CTMM).<sup>1</sup> These "noncritical" items were sorted by four judges into three difficulty categories (easy, medium, and difficult), and from these sorts three basic test forms (easy, medium, and difficult) having 54 item positions, of which 45 contained "noncritical" items, were constructed. Every sixth item position was left blank in the basic test forms.

The nine "critical" items, all of which were considered by the judges to be very difficult, were taken from the CTMM and other sources. The stems of these "critical" items were bowdlerize, detritus, obloquy, orrery, philippic, putative, quiddity, syzygy, and termagant. These nine items were arranged in a balanced order across positions and tests in the blank-sixth positions (positions 6, 12, 18, 24, 30, 36, 42, 48, and 54) of the easy, medium, and difficult basic tests. Each of the nine items appeared in the first third (either position 6, 12, or 18), second third (either position 24, 30, or 36), and last third (either position 42, 48, or 54) of the easy, medium, and difficult tests.<sup>2</sup> Such an arrangement necessitated the use of three forms of each of the three basic tests, or nine test forms in all.

The nine tests were administered to 297 women students (33 students per test) in an introductory psychology course at a southeastern college. The tests were distributed in an alternating order within several course sections, so that approximately equal numbers of students in a given section took each of the nine test forms. Directions to the students specified that there were no special instructions about guessing on difficult items, but they were told that they should try to get as many items right and as few wrong as possible. Although the directions stated that there was no time limit, no student took more than 25 minutes to finish the test. The students were not informed that three tests of different difficulty levels were being administered simultaneously.

### C. RESULTS

The completed test papers were scored according to the number of items correct and, at each difficulty level, the median scores for the three forms combined were determined. The median scores for the easy, medium, and difficult tests were 46.38, 39.65, and 20.10, respectively. Because the median scores of the easy and medium tests were close in value and both were different from that of the difficult test, the first two groups of scores were treated as a

<sup>1</sup> The writer wishes to express his appreciation to the California Test Bureau for permission to reproduce these items for use in this investigation.

unit in subsequent analyses. One noteworthy fact is that on only 66 of the 297 papers was there at least one "critical" item left blank.

The following are the results of the chi-square test of the independence of the variables of test difficulty and tendency to leave "critical" items blank. The data format for the chi-square test was a 2-by-2 contingency table of the numbers of papers having one or more of the nine "critical" items blank *vs.* the number having no "critical" item blank as one dichotomized variable, and the easy and medium tests combined *vs.* the difficult tests as the other dichotomized variable. The result of the chi-square test ( $\chi^2 = 5.41$ ,  $df = 1$ ,  $p < .02$ ) showed that a significantly greater per cent of the difficult papers than of the easy and medium papers combined had at least one "critical" item blank. A further analysis of the 66 papers having at least one "critical" item blank showed that the per cent of the difficult test papers also having at least one "noncritical" item blank was greater than the per cent of easy and medium tests combined also having at least one "noncritical" item blank ( $\chi^2 = 31.97$ ,  $df = 1$ ,  $p < .001$ ).

Finally, an analysis of the positions of the blank "critical" items in the 66 papers having at least one "critical" item blank showed that, for all 66 papers combined, the total number of "critical" items left blank was not significantly different from expectation in the first, second, or last thirds of the test ( $\chi^2 = 4.30$ ,  $df = 2$ ,  $p < .20$ ).

#### D. DISCUSSION

These results support the hypothesis that the difficulty context in which multiple-choice test items are placed has an effect on whether or not the items are attempted. It should be emphasized that the use of the number of "critical" items left blank on a nontimed, multiple-choice vocabulary test is a stringent criterion of this effect since less than one-fourth of the examinees left one or more "critical" items blank. Therefore, the fact that significantly more than expected of the tests with at least one "critical" item blank were difficult tests would appear to be good evidence for a difficulty-context effect.<sup>2</sup>

<sup>2</sup> A similar effect of difficulty context on leaving critical items blank has been found by the writer for short-answer, no-time-limit, mathematics tests. One of six mathematics tests (two forms each of easy, medium, and difficult tests) of 24 items, six very difficult items being common to all tests, was administered to each of six groups of 26 students. An analysis of the number of critical items left blank on the easy, medium, and difficult tests gave a significant overall chi square ( $\chi^2 = 15.37$ ,  $df = 2$ ,  $p < .001$ ). Analysis of the number of critical items left blank for the easy *vs.* medium, easy *vs.* difficult, and medium *vs.* difficult comparisons gave chi squares of 5.96, 15.24, and 2.88, respectively. Since  $df = 1$  for each of these comparisons, the first two chi squares are statistically significant, but the last is not.

The finding that, among the 66 papers with "critical" items blank, the difficult tests were more likely than the easy and medium tests also to have "noncritical" items blank is open to a number of interpretations. It may be argued that leaving certain items blank induces an "OK set" to leave other items blank. Another possible explanation is that the high degree of negative reinforcement associated with a more difficult test causes the examinee to become less interested in trying difficult "critical" items. This kind of reinforcement interpretation has been used by Carr and Sax (1) in their related study.

Finally the result that, in the 66 tests with at least one "critical" item blank, the total number of "critical" items left blank is similar in the first, second, and last thirds of the test is consistent with the results of previous research on item-position effects.

### E. SUMMARY

The effects of difficulty context and item position on tendency to leave items blank were studied with three different forms of each of three multiple-choice vocabulary tests of varying difficulty levels (easy, medium, and difficult). It was found that the tendency to leave blank at least one of the nine very difficult "critical" items included in each test was associated with the difficulty level of the test. Of the 297 completed test papers, the per cent of difficult tests having one or more "critical" items blank was greater than the per cent of easy and medium tests combined having one or more "critical" items blank. On the other hand, the results did not support the hypothesis of a significant item-position effect. Of the 66 papers with at least one "critical" item left blank, the total number of "critical" items left blank was not significantly different from expectation in the first, second, or last thirds of the test. These findings are consistent with the results of previous studies of the effects of item context and position on responses to multiple-choice test items.

### REFERENCES

1. CARR, A., & SAX, G. An investigation of response set on altered parallel forms. *Educ. & Psychol. Meas.*, 1962, **22**, 371-376.
2. CRONBACH, L. J. Response sets and test validity. *Educ. & Psychol. Meas.*, 1946, **6**, 475-494.
3. ———. Further evidence on response sets and test designs. *Educ. & Psychol. Meas.*, 1950, **10**, 3-31.
4. MARCUS, A. The effect of correct response location on the difficulty level of multiple-choice questions. *J. Appl. Psychol.*, 1963, **47**, 48-51.
5. MOLLENKOPF, W. G. An experimental study of the effects on item-analysis data of changing item placement and test time limit. *Psychometrika*, 1950, **15**, 291-315.



## AESTHETIC PREFERENCES OF ART STUDENTS AND MATHEMATICS STUDENTS\*

*Psychology Department, University of Georgia*

---

RUSSELL EISENMAN AND SANDRA COFFEE

---

### A. INTRODUCTION<sup>0</sup>

This experiment represents an extension of the findings of Taylor and Eisenman (5), who found that creative art students preferred complex polygons, in contrast to less-creative art students who more frequently chose simpler polygons as their most preferred choices. Taylor and Eisenman used polygons from the symmetrical forms of Birkhoff (2) and the asymmetrical forms of Vanderplas and Garvin (6). Because they were interested in determining the amount of complexity preferred by their creative and less-creative art students, these investigators employed polygons of varied complexity, as defined by the number of points on the forms.

Martin<sup>1</sup> recently questioned the notion that artists would prefer more complex forms. The whole question of what is aesthetically preferred has been a disputed one. Some have taken what may be termed an objectivist stand, claiming that in objects there are inherent qualities that make for aesthetic beauty. The opposite viewpoint emphasizes that the aesthetic value of objects is a function of the values of the person doing the observing. An objectivist view was provided by the Harvard mathematician Birkhoff, who felt he had devised a formula for determining aesthetic preference (2). The failure of psychological studies to support Birkhoff's formula is documented by Eisenman (3), who has shown that, while there has not been total rejection of Birkhoff's formula—a few studies provide partial support—Birkhoff's formula of aesthetic measure fares poorly.

Birkhoff's attempt to devise a formula for aesthetic measure in general led him to conclude that simpler forms are most preferred. The present study indicates that Birkhoff's finding of a formula that indicated simpler polygons were more aesthetically pleasing may have been related to his being a mathematician rather than an artist. At any rate, the present study was undertaken to see if art students and mathematics students would prefer different kinds of polygons.

---

\* Received in the Editorial Office on June 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Personal communication, May 15, 1964.



## B. METHOD

### 1. Subjects

Twenty college seniors served as *Ss*. Ten were majoring in art, and 10 were majoring in mathematics. The *Ss* appeared to be ignorant of the purposes of this study, as might be expected. Several of them asked whether or not the study was a personality investigation, equating looking at polygons with taking a Rorschach test.

### 2. Materials

The stimulus objects consisted of 10 polygons. Five symmetrical forms were selected from the 90 figures of Birkhoff (2), and five asymmetrical forms were selected from the 180 figures of Vanderplas and Garvin (6). All 10 polygons had six points, thereby keeping complexity constant. The polygons were photographed and enlarged to 4"  $\times$  5" prints by a commercial studio.

### 3. Procedure

*S* sat at a table facing *E* and was given a piece of paper and a pen with which to record his preferences. Listed on the paper were the numbers one through 10. *S* was then told:

I am going to give you 10 cards. I want you to rank them in order of your aesthetic preference. You are to pick out in order your first most preferred, second most preferred, third, and so forth until you come to your tenth most preferred. These are to be listed on the paper provided. Take your time. If two appear to be a tie as far as you are concerned, choose one or the other. Do not rank them the same. When you have completed this, do not change any of your answers, and hand them to me.

The number of the card was marked on the back at the top right-hand side. These numbers were Roman numerals because the paper on which *S* marked his preferences contained Arabic numbers. It was felt this distinction would make the task clearer, avoiding chance errors.

*E* tried to avoid questions as much as possible once the experiment had begun. Most questions related not to the procedure employed, but to whether or not the experiment would reveal something of *S*'s personality. These *Ss* were informed that this was not the purpose of the experiment.

## C. RESULTS AND DISCUSSION

Table 1 presents the first choices of the 10 art students and the 10 mathematics students.

It is apparent that the hypothesis of different preferences by these two groups is strikingly supported. Every single one of the art students chose an asymmetrical polygon as his first choice; every single one of the mathematics students chose a symmetrical polygon as his first choice ( $\chi^2 = 20.00$ ,  $p < .001$ ). For nine of the 10 polygons, the chi-square test indicated a difference significant beyond the .05 or .025 levels.

The results contradict any hypothesis that implies that there are objective aesthetic preferences for mankind as a whole. The art students preferred the asymmetrical polygons, while the mathematics students preferred the symmetrical ones. If we think of the asymmetrical forms as the more complex,

TABLE 1  
FIRST CHOICE ON POLYGONS BY ART AND MATHEMATICS STUDENTS

Students	Polygons	
	Symmetrical	Asymmetrical
Art majors	0	10
Mathematics majors	10	0

then this experiment supports the contention that artists prefer more complex forms.

What is left of the popular belief that artists prefer such qualities as simplicity, order, and balance? The results of this experiment do not demolish such a belief, though these data should cause one to examine critically such a notion. A possible explanation is that the more complex forms—which artists prefer in this study and in studies by Taylor and Eisenman (5), Barron and Welsh (1), and Eysenck (4)—are really not so complex to the artist.<sup>2</sup> The artist may be able to see a great amount of order in what appears to the non-artist to be highly disordered and of an indisputably complex nature.

Such a point may have been entirely overlooked by Birkhoff (2) when he devised his formula. Because Birkhoff was a mathematician he may have been biased in favor of simpler forms, and this bias might have spread to the formula that he derived.

The failure of Birkhoff's formula to stand the test of time may also be ascribed to his philosophical position, which was apparently one of objectivism. At least this is the philosophical position inherent in a formula that is supposed to represent the aesthetic preferences of everyone. A more prudent approach would be to suspect that different kinds of groups prefer different sorts of things, and that such preferences may change as a function of changing

<sup>2</sup> This point grew out of a stimulating discussion with Robert E. Taylor, to whom we are thankful.

historical events. Thus, in the industrialized Western world of today, it may well be that people in general tend to prefer more complex forms than people in a more agrarian community. But, if this were found to be the case, the mistake should not be made of considering it an immutable truth.

#### D. SUMMARY AND CONCLUSIONS

This experiment was conducted to see whether art students and mathematics students would prefer different kinds of polygons: symmetrical *vs.* asymmetrical. The results indicate that the art students preferred the asymmetrical forms, while the mathematics students preferred the symmetrical polygons.

The results of this experiment are inconsistent with an objectivist viewpoint which maintains that there is something inherent within objects that makes them universally beautiful. Such a philosophical belief may have been part of Birkhoff's thinking when he devised a formula of aesthetic measure. This possibility may account for the lack of support which such a formula has received from various researchers.

#### REFERENCES

1. BARRON, F., & WELSH, G. S. Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. *J. of Psychol.*, 1952, **33**, 199-203.
2. BIRKHOFF, G. D. *Aesthetic Measure*. Cambridge: Harvard Univ. Press, 1933.
3. EISENMAN, R. Creativity, complexity, and meaningfulness in visual art. Unpublished Master of Science thesis, University of Georgia, Athens, 1963.
4. EYSENCK, H. J. The empirical determination of an aesthetic formula. *Psychol. Rev.*, 1941, **48**, 83-92.
5. TAYLOR, R. E., & EISENMAN, R. Perception and production of complexity by creative art students. *J. of Psychol.*, 1964, **57**, 239-242.
6. VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. *J. Exper. Psychol.*, 1959, **57**, 147-154.

*Psychology Department*  
*Milledgeville State Hospital*  
*Milledgeville, Georgia*

## DESCRIPTION OF A CENTRAL MECHANISM RELATED TO DIVERSE BRIGHTNESS PHENOMENA\*<sup>1</sup>

*Department of Psychology, Michigan State University*

---

THOMAS M. NELSON AND S. HOWARD BARTLEY

---

### A. INTRODUCTION

The ideal to approach in dealing with relations between photic input and sensory end-result is as complete an understanding of the intervening body processes as possible. It has long been assumed that the peripheral mechanism, the sense organ, plays the role of defining the kinds of discrimination the organism can make. Helmholtz expressed this belief by saying, "It is sufficient to assume that the optic nerve is capable of sensations of different kinds without trying to find out why the system of the visual sensations is just what it is" (24). Thus the differential reactions the organism can make to wavelengths of photic input (*viz.*, seeing color) have been supposed to be determined by the eye. That is to say, the cerebral cortex is not supposed to make differentiations but only to use those given it. Accordingly the study of color, brightness, etc., has centered on what happens in the eye. Only recently, with the advent of microelectrode recording has a detailed sampling analysis of the cell activity of the central end of the optic pathway been made. This recording has been more on the nature of a cell taxonomy than on the production of an action picture of the system. It has also been a kind of carrying through of a circuitry picture from a given patch in the retina to given cells in the cortex, still not a full-fledged action description of the *role* of the cortex.

There are several loci in the optic pathway to investigate, and each will contribute to an ultimate understanding of the physiological machinery for vision. But it must not be assumed that understanding the more central segments of the pathway awaits the full description of how the periphery (sense organ) works. It is as reasonable to eliminate the sense organ momentarily to find out how the central end of the pathway works as it is to eliminate the central portion of the pathway to study the sense organ and the sensory end-

---

\* Received in the Editorial Office on June 24, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The present paper is a part of the work being done under Research Grant NSF-G-19485 from the National Science Foundation.

result. And it was just such experimentation that permitted the time-intensity conceptualization of the visual pathway known as the *alternation-of-response theory*.

### B. PROBLEM

Stimulus timing is critical for the production of many brightness phenomena. Specific cases are expressed in the Bunsen-Roscoe law, Block's law, Talbot-Plateau law, Blondel-Rey curve, Broca-Sulzer curve, and brightness enhancement or Brücke-Bartley phenomenon. Each one of these psychophysical laws or manifestations is demonstrable under the conditions disclosing it, but when all of them are examined in relation to each other certain limitations or incompatibilities are apparent. For example, Baumgardt (15) showed time-intensity reciprocity (Bunsen-Roscoe law) to hold for pulses as long as  $1 \times 10^{-1}$  second, while Stewart (32), Beams (16), and Gilmer (23) showed average flux sufficient to predict brightness (i.e., the Talbot-Plateau law to hold) when repeated pulses were no longer than  $8 \times 10^{-7}$ ,  $1 \times 10^{-8}$ , and  $8 \times 10^{-9}$  second respectively. Also a stimulus field made to flicker by proper cycling possesses a brightness in excess of the same field both when it is nonintermittent or made intermittent rapidly enough to produce fusion (i.e., above critical flicker frequency). This is the brightness-enhancement phenomena. It would seem that possibly the conditions under which these phenomena appear are more restricted than common views have interpreted them to be.

This paper will interrelate these phenomena through the mechanism described in the alternation-of-response theory. It must be understood that this theory is not a fixed statement—a psychophysical formulation—meant to substitute for the others by combining them, but a description of what the central end of the pathway essentially consists in and how its elements are temporally related in their activity. The theory refers chiefly to the way input to the cortex is handled and not to descriptive correspondence between sensory end-result and retinal events. The assumption underlying this approach is that all the limitations and possibilities of the visual system are represented at or located within the final neural stage just preceding brightness discrimination. This, however, can only be taken to mean that the mechanism outlined in the alternation-of-response theory is a necessary component in understanding the facts summarized by the laws, phenomena, curves, and effects depending upon stimulus timing. It is not a total account since the initial event (i.e., the eye) has been left out of consideration. The fact that enables one to leave the initial event out of consideration rests upon the original finding that certain broad central phenomena were essentially the same for the analytical purposes



used, regardless of whether the input (electrical) was delivered to the stump of the optic nerve or (photic) to the retina.

### C. ALTERNATION-OF-RESPONSE THEORY

It was in the interest of studying central function that Bishop and Bartley for many years studied the cortical response to controlled electrical stimulation of the optic nerve (eye removed). The study yielded a group of dovetailed results which enabled a very workable conception of central activity. While it left out (did not ascertain) many of the fine details, it did depict a well-established general framework of activity useful as a neurophysiological understanding and provided an understanding that has also proved useful in dealing with relations between photic stimulation of the eye and sensory end-results in human-perception experiments.

The essential feature in bridging the gap between artificial (electrical) and natural (photic) inputs was the fact that the brief stimuli of both sorts that these investigators used produced essentially the same character of cortical response. This correspondence was all that was needed to infer that the general principles of response activity being recorded at the cortex were those not obscured by the details of retinal elaboration. It may be difficult to envisage how this could be when so many features of ultimate sensory response are traditionally supposed to be fundamentally determined in the eye—nevertheless, it seems appropriate to accept the facts of the case and to suppose that nothing noted in the cortical response under the Bishop-Bartley experimental conditions contradicts the true role played by the eye. It is to be supposed that the eye provides many of the crucial fine details that elaborate upon the general temporal intensity pattern that Bishop and Bartley disclosed.

The purpose of the foregoing introduction of the general neurophysiology of the optic pathway has been to bring in what seems to be the kind of information necessary to account for the relative brightness produced by intermittent and steady stimulation. No photochemical feature of receptor function has as yet come anywhere near to accounting for the complex of sensory end-results expressed in the experimental data from a variety of investigators.

The alternation-of-response theory provides a time-intensity picture of the central end of the optic pathway. It consists of a set of statements regarding the way the optic pathway responds in relation to the quantitative (intensive, durational, and distributional) aspects of photic input. It predicts what is to be expected in cortical response and sensory end-results when various features of the pathway activity are altered by known forms of input.

The theory explains the major and essential features—not the fine details which typically concern so many workers. It is the very picture that seems to deal best with the facts of brightness response. Many aspects of the alternation-of-response theory have been published. For this reason it may seem superfluous to repeat them; however, there are several features that need to be explicitly developed for present purposes. These features are best understood in the context of some of the frequently reiterated items of the theory (2, 4, 5, 6, 9, 10, 11, 12). It involves the following combinations of facts and suppositions.

The number of parallel channels in an animal's optic pathway is constant. In part of the pathway, these channels are the optic-nerve fibers. The channels cannot be activated continuously. Response alternates with recovery. The response-recovery cycle is about one-fifth second in the rabbit. In the human, it is around one-tenth second. Either photic or electrical stimuli may activate a varying number of channels, depending upon input strength and duration and rate of repetition. Fairly brief intense inputs activate all or nearly all the channels of the pathway together.

The effectiveness of subsequent inputs depends upon their temporal relation to the first. The channels having been activated together by the initial input tend to recover together, though recovery time differs from channel to channel. The greatest opportunity to reactivate all channels together results when brief isolated inputs are used and input rates are about 10 cps or below. Proper timing of inputs in a series can keep a maximum number of channels in synchrony with one another although amplitude of cortical response will not be as great as that produced by isolated inputs.

If input is continuous rather than repetitive (intermittent), there will be as many channels in action as are recovering at any and all instants. In other words, uniform, prolonged inputs will elicit uniform overall response of the pathway. Although, with uniform prolonged inputs, the overall activity of the pathway is uniform, the uniformity does not apply to single channels. These rotate in their activity, as first indicated. For example, when a second input is presented before all of the channels originally activated by the first input have recovered, the second input will activate only those that have recovered. A third input will find still other channels recovering or already fully recovered and will be able to activate them. Actually, the third input may be delivered at such a time as to reactivate most of the channels earlier activated by the first. Thus, alternation-of-response reaches its limits when inputs become uniform and indefinitely prolonged. The first brief portion of the prolonged input sets a number of channels into action. If the impingement is maximal, the number of channels activated will be maximal. If not, some lesser number will

be activated. If the maximal number is activated simultaneously, no more activation can occur until some recovery has taken place. Sooner or later a uniform rotation in the activation and coming to rest of the various channels will be accomplished. This is one of the commonest forms of neural activity. It applies not only in the visual but in the tactual (22), the kinesthetic (31), and possibly the auditory (33) system as well.

Under many circumstances, there is a parallel between the size of the cortical response, and the sensory end-result. For example, factors tending to produce cortical responses of high amplitude likewise result in high brightness. This was the first fact found to show correlation between a specific item of neurophysiology and sensation. No continuous input can be as effective per unit of time in producing brightness as a very brief input (other variables being equal). Brief photic pulses should be more effective in producing brightness than longer ones, for long pulses function as "continuous stimulation" which repeatedly reactivates channels.

The possibility that not all conditions in the optic pathway related to the emergence of the experience of brightness may be measurable as something occurring at a single region in the cortex must be recognized. To the extent that brightness is a function of the pattern of activity of a more widespread tissue area or volume, the theory is open to refinement.

#### D. OPERATIONAL EQUIVALENTS

The specific neurophysiological phenomena from which the alternation-of-response theory was derived were represented as neurophysiological records. One may summarize the theory on an operational level as well.

The chief cortical variables related to brightness response by the present theory are amplitude-of-response components and density-of-channel discharges per unit of time. The two variables are interrelated. If a given number of channels discharge synchronously, one can expect amplitude in the recorded cortical response to increase. If the channels are desynchronized, one can expect density to increase. Given a sheer increase in the total number of operating channels, one may expect both amplitude and density to increase since the total activity per unit of time is greater. Partly responsible is the incidental increase in diversity of recovery times and latencies of the individual channels. With increase in the total number of channels discharging at a given degree of synchrony, only amplitude is increased; whereas the same increase in total number of channels active would increase density if evenly spread over time. If channel discharges are grouped at rates slower than that defined by the recovery time of a single channel, one may expect amplitude to

increase. All discharges occurring at rates less than around 10 per second tend to maximize synchrony and minimize density (11).

The sensory states produced from intermittent photic input are basically those of flicker and fusion. The point of transition between these two states is a variable that is precisely measurable for a specified set of conditions and has received extensive study. Since the sensory end-result at critical flicker frequency (CFF) has not been shown to change with further increases in rates, conclusions concerning the condition of the visual mechanism at CFF pertain as well to the whole fusion range. The chief facts of CFF or fusion are as follows: (a) CFF varies with intensity—when unilluminated surrounds are used, CFF increases with intensity to a point and then declines; (b) CFF varies with pulse-to-cycle fraction (PCF) for unmodulated stimulation—it increases with PCF to a point and then declines; and (c) CFF decreases as the separating interval between short pulse trains decreases.

Flicker, unlike fusion, is a highly variable phenomenon (5, 8). Stages of glitter, flutter, coarse flicker, and marginal flicker can be observed to accompany changes in rate, PCF, and intensity. In addition there is a considerable range of other sensory effects associated with the progressive changes of flicker appearance. Brightness is not fixed below CFF as it is above CFF, but varies as different rates of input are combined with variables of (a), (b), and (c). See Bartley (3). The minimum angle of resolution is increased for structured fields (14). Full-spectrum targets develop specific hue of some saturation when various rate-PCF combinations are employed (21), and part-spectrum targets are grossly altered in hue and saturation (1, 25, 27).

It is clear that the variability of sensory detail is much richer when inputs produce flicker than when they produce fusion. Adequate analysis of flicker is apparently a requisite to a sound explanation of fusion. The present discussion, for this reason, will focus attention upon the mechanism maintaining flicker rather than upon that for producing fusion: i.e., the mechanism maintaining flicker when conditions are tending toward CFF rather than upon the mechanism whereby photic response is "carried over the dark interval."

Density is a time-dependent variable related to pathway activity. It is associated with variations in synchrony-asynchrony without being simply reducible to synchrony-asynchrony. Density refers to the discharge activity per unit time and is defined as the reciprocal of separations between channel discharges. The average level of density falls when intervals between successive channel discharges are uniformly great. The spacing between successive discharges can increase either because of a decrease in input rate or because of a reduction



in the number of channels in action. However, average density is not to be directly associated with sensory end-results of flicker and fusion.

Flicker and fusion seem associated with uniformity of discharge density. At low rates of repetition of photic input, average channel activity is very dense over a brief period. But this is followed by a period in which average density is very low, the effect being a highly variable density. As rate of input repetition is raised, discharge becomes increasingly uniform since the channels in synchrony become fewer and fewer, and conditions for simultaneous discharge of channels become progressively less effective. Near CFF, most channels are firing at their own unique maximum rate. Our hypothesis, therefore, is that *flicker persists so long as critical values in irregularities of density (and necessarily amplitude) persist.*

The same *average density* may exist with different types of temporal distribution of channel activity. That is, amplitude may fluctuate greatly in one case and only slightly in another. Different average densities may exist when the periodicity of fluctuation is the same. An increase in average density increases brightness only when periodicity of fluctuation is held constant.

It follows from this that flicker and brightness need not always covary in any simple way. Brightness will increase to the extent that sheer amplitude of response from instant to instant varies while average density holds constant. Brightness will undergo a continued but not necessarily constant increase as any given input duration is separated from that preceding it until maximum synchrony is achieved. Brief inputs are seen as isolated and reach maximum brightness when uniformity of density is so poor over the cycle as to be said to be absent. At this point, repetitive stimulation results in a sensory impression of successive single flashes instead of a continuity possessing a fluctuating brightness (flicker).

#### E. THE TALBOT-PLATEAU LAW AND THE BRIGHTNESS OF RESTRICTED NUMBERS OF PHOTIC INPUTS

An investigation critical to the development of the theory is in press (26). In this investigation, CFFs of short and long durations of intermittent stimulation, such as indicated in Figure 1, were presented to the eye. CFFs were established for each and then these same trains were matched for brightness at the CFF rate. The brightness, flicker, fusion, and CFF levels are described in Figures 2, 3, and 4.

The upper panels of Figure 2 indicate the brightness indexes for pulse trains of various lengths, interaction characteristics, and intensity. The result is that nearly all points are above the Talbot-Plateau level.



## SAMPLE PHOTIC PULSE SEQUENCES



FIGURE 1

SCHEMA SHOWING SOME OF THE TEMPORAL DISTRIBUTIONS OF PHOTIC INPUT USED IN EXPERIMENTS TO DETERMINE FACTORS INFLUENCING BRIGHTNESS AT CRITICAL FLICKER FREQUENCY

The variables are numbers of pulses in a train, separation between trains, pulse intensity, and pulse-to-cycle fraction (PCF).

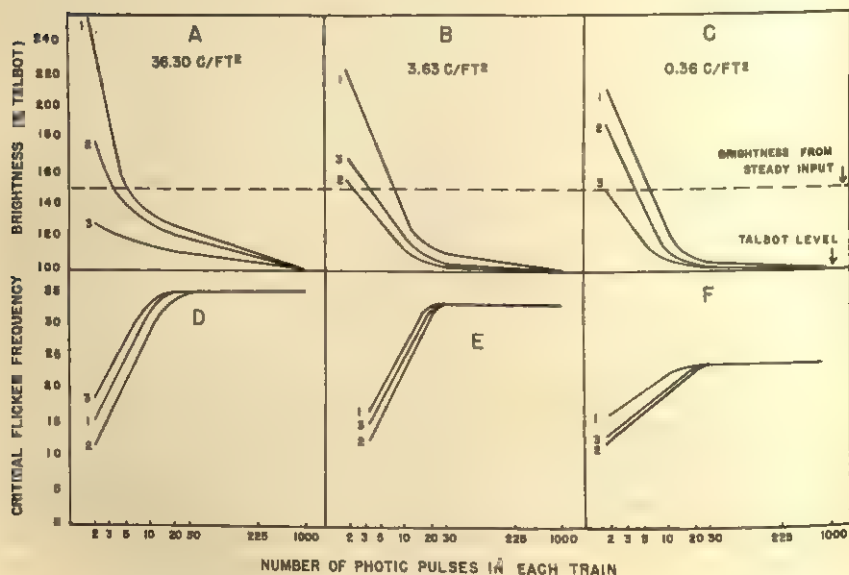


FIGURE 2

RELATION BETWEEN NUMBER OF PHOTIC PULSES IN EACH TRAIN, BRIGHTNESS, AND CRITICAL FLICKER FREQUENCY (CFF)

Columns, from left to right, show descending levels of intensity. Curves labeled 1 are for trains so widely separated as not to involve interaction (PCF .67). Curves labeled 2 and 3 are the results with small separations, three and six cycles respectively, in which interaction occurred.

The lower panels of Figure 2 show that, as the number of pulses in the trains increase, the maximum separation must be reduced to maintain CFF. This tallies with previous findings (13).

In the upper part of Figure 3, the two response indexes (brightness and CFF) are plotted against each other. At the two lower intensities, the opposition in order of values is virtually linear until CFF becomes maximum. When CFF is maximal, however, brightness is still above the Talbot-Plateau level, but declines to that level as train length is increased.

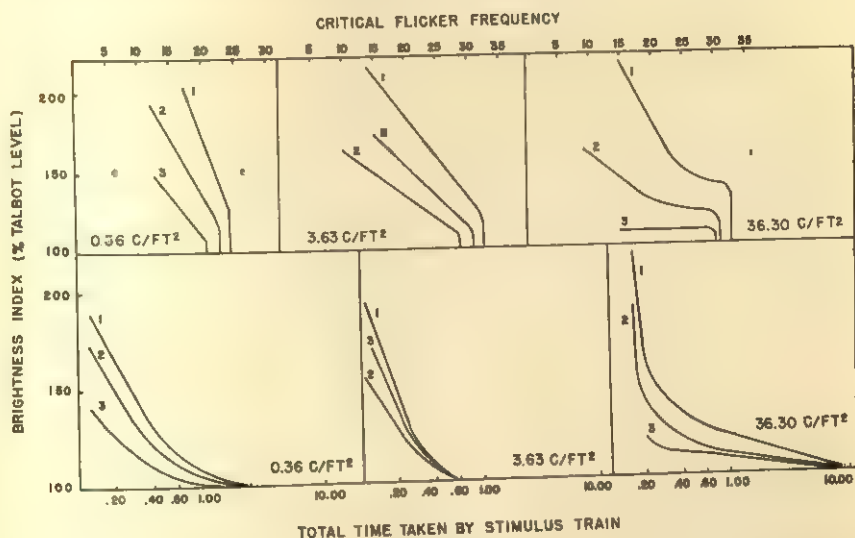


FIGURE 3

THE RELATION BETWEEN BRIGHTNESS IN TERMS OF THE PER CENT OF THE TALBOT LEVEL AND CRITICAL FLICKER FREQUENCY FOR EACH OF THREE LEVELS OF INTENSITY (UPPER ROW) AND THE RELATION BETWEEN TOTAL TIME TAKEN BY THE SHORT TRAINS AND BRIGHTNESS AT CFF

The curves labeled 1 depict wide separations that would avoid train interaction; and Curves 2 and 3, separations of six and three cycles respectively.

The lower curves of Figure 3 give the relationship between the total time period occupied by the fused-pulse train and brightness. Brightness declines as the duration of fused presentation increases. The curves are negatively accelerated in all cases and appear to have the form reported by Broca and Sulzer (20) for single photic inputs well above threshold. In general, the brightness produced by trains that do not seem to interact is greater than for those that follow each other so closely as to interact, for equally long periods of presentation producing fusion.

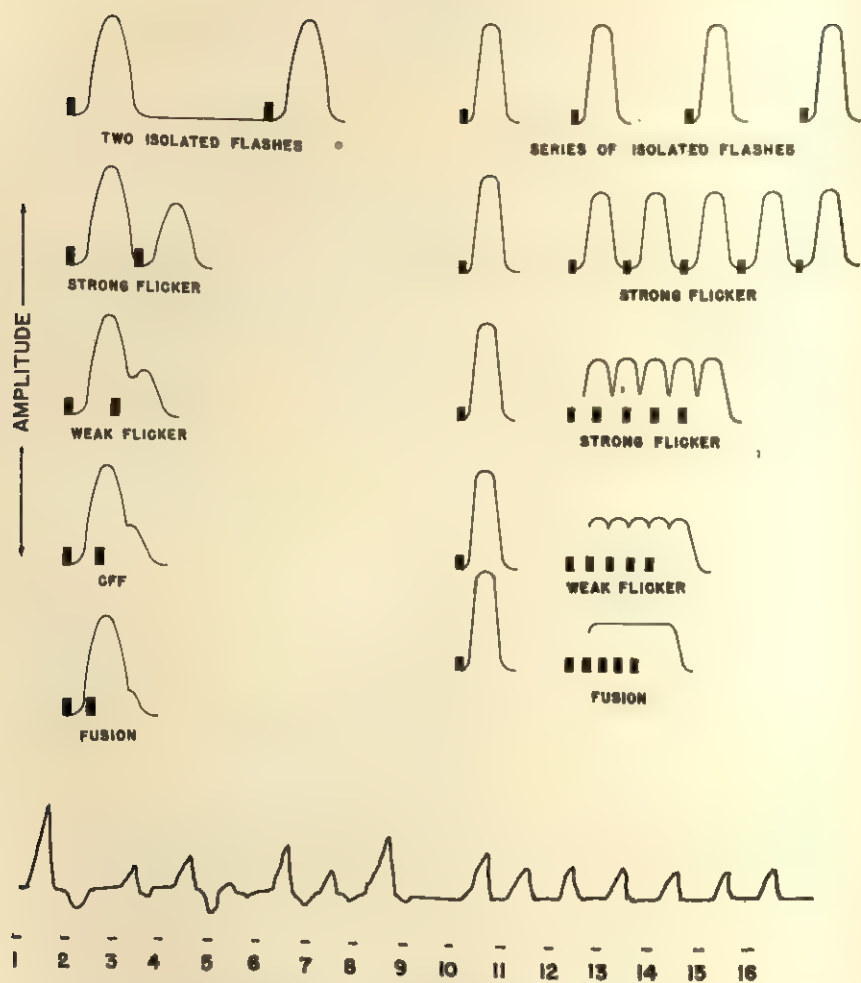


FIGURE 4

SCHEMATIZATION OF THE CORTICAL AND SENSORY CORRELATES OF PROGRESSIVELY VARYING THE SEPARATIONS BETWEEN PULSES

The first column depicts the schemata for two pulses; the second, for a greater number of pulses. See text for discussion.

## F. INTERPRETATION OF EFFECTS FROM SHORT DURATION OF INTERMITTENT STIMULATION

According to the older view (Ferry-Porter law), decreases in brightness are not expected to be associated with increase in CFF. Nevertheless this relationship has been found to exist under conditions such as those resulting in Figures 2 and 3. The neurophysiological view presented accounts for it because it is presumed that concurrent changes in density and amplitude are involved in the two end-results.

As the number of closely spaced pulses in a train is reduced and the separating intervals between trains are held constant at some value larger than the period between pulses in the train itself, response amplitude tends to increase. The first member of any photic train is likely to find a maximum number of channels in readiness and able to be activated simultaneously (7). The second pulse activates only those channels having recovered during the period separating the first from the second members. When a train consists of only two pulses, the first pulse activates all or nearly all the available channels. The second pulse will not be equally effective until recovery of channel activity to the first pulse is complete. With long-pulse trains, the alternation-of-response process sets in, whereby the pulses can be delivered closer and closer together and still maintain flicker. Concomitant with this, the various changes in amplitude and density are occurring whereby brightness is made to decline.

The recovery of the channel mechanism is known to be slower than that of the retinal mechanism (19), so that the temporal determination of the results cannot be mainly attributed to the retina. Bishop's work (17) indicates that the separation between members of a stimulus pair must not be less than 100 msec for the average observer if two distinct cortical discharges are to be produced. The cortical and sensory correlates of progressively varying the separation between two pulses of fixed duration is schematized in the first column of Figure 4 and, for a longer train, in the second column. Differences in activation and recovery time among channels accumulate as the number of pulses in the train is extended. This might seem to work toward greater asynchrony and thereby reduce CFF. However, the fact that CFF has been shown to increase indicates that something in addition to sheer increase in asynchrony must account for sensory end-result. The increase in average density affords an increasing likelihood of the more frequent inputs activating certain groups of channels in synchrony and would seem to account for the sensory phenomena obtained. This is to say that, while longer trains of channel stimulation work progressively toward more uniform distribution of activity,

the probability of synchronous discharge of small numbers of channels also rises. This "residual synchrony" is reflected in the appearance of the target field. The long-train flicker just preceding fusion is very much less coarse than short-train flicker. In the latter, the whole target seems involved and only very marginal or superficial effects are involved in the former. Before fusion may be reached, residual synchrony must be obliterated. For this to occur, density must be increased to a critical point.

It was pointed out earlier that the curves for the lowest intensity condition of Figure 3 (upper part) are the steepest. Presumably this occurs because the number of channels activated is fewer, and hence a more limited distribution of latencies and recovery times is involved than at higher intensities. As intensity is increased from .36 c/ft<sup>2</sup> to 3.6 c/ft<sup>2</sup> to 36.3 c/ft<sup>2</sup>, the absolute number of channels in action increases and, because of this, any given rate of stimulation brings about cortical discharge of greater amplitude and density. The change from full synchrony to asynchrony therefore provides for a greater loss of amplitude and gain in density. This is reflected in Figure 3 (upper part) both as an increase in the range of conditions over which the brightness index varies and as an increase in the range of CFF measures as intensity is increased.

In Figure 3 (upper part), the increase in the range of rates producing CFF is considerably larger than the range of increase in brightness. This suggests that amplitude of discharge is less affected by intensity increases than is density. If amplitude undergoes less change because stimulation rises above threshold for more and more channels, rates of firing of the channels in total become more heterogeneous and hence not so likely to synchronize fully. This heterogeneity in channel activity would result in some synchrony being maintained even when periods of stimulation are short; thus, as the highest intensity curve of Figure 3 (upper part) indicates, input rates can be quite high and still support flicker.

The slopes of CFF-brightness curves are always less steep for pulse trains so closely separated as to affect each other than are those that are farther apart but which have the same number of repetitions. The separating interval is so short when CFF is reached for the pulse trains themselves that many of the channels still have not recovered when the first pulse of the succeeding train appears. Under identical conditions of rate, lower amplitudes are produced and uniform densities are more easily obtained. This depresses CFF.

Figure 5 depicts the relationship between brightness and period of stimulation expected for pulse trains possessing varying numbers of repetitions. Left



to right, the schema depicts the passage from intermittency rates that are so rapid as to collapse the entire pulse train into a duration lying within the Bunson-Roscoe period to intermittency rates so slow that each pulse appears independent of other inputs preceding and following it. The Bunson-Roscoe portion of Figure 5 shows time-intensity reciprocity; the brightness in all

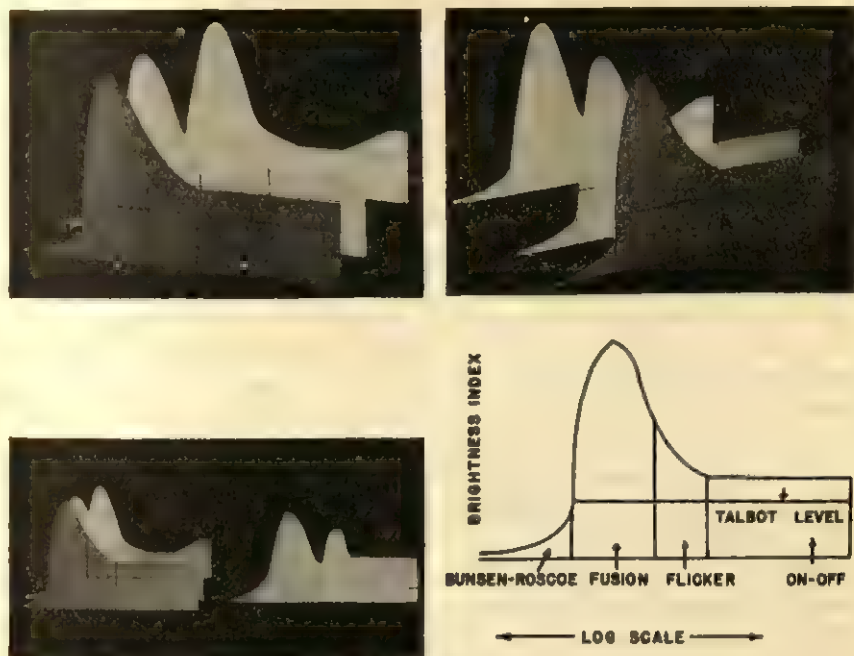


FIGURE 5

SCHEMA FOR RELATIONSHIP BETWEEN BRIGHTNESS AND PERIOD OF STIMULATION  
FOR PULSE TRAINS OF VARYING NUMBERS OF REPETITIONS

Time is depicted on the horizontal axis, brightness on the vertical. The schema shows that, at CFF, brightness varies above and below the Talbot level.

cases is therefore necessarily below that predicted for fused stimulation by the Talbot-Plateau formula. With intensity constant, the number of peripheral elements responding or their rate of firing varies in proportion to the time occupied by the train. This reduces amplitude of cortical discharge even though synchrony is virtually perfect.

The portion of Figure 5 following the Bunson-Roscoe period shows brightness variations when rate is rapid enough to produce fusion but longer than the times producing reciprocity. The conditions used in the study just described fall within this range. The curve approximates the Broca-Sulzer form. The

peak of the curve reproduces the amplitude occurring with virtually synchronous discharges of all available elements (those capable of discharges under the given intensity). Were the modeled intensity very low, a Broca-Sulzer increase would not have been shown; but rather the curve would gradually approach Talbot-Plateau brightness level in the manner expressed by the linear relationship of Blondel and Rey (18). Consistent with our data, only long trains reach the Talbot-Plateau level. The low point of brightness of the other pulse trains always equals or exceeds the Talbot-Plateau level. That of the two pulse trains in fact never drops below the level produced by steady illumination.

Estimations of brightness levels below CFF are also given in accord with prior findings (28, 29, 30). There is a gradual change to steady illumination beginning at a rate of three cps for the pulse-to-cycle fraction (PCF) schematized. Pulses at that rate are 222-msec long and are separated by 111 msec. With these durations, single pulses provide the same amplitudes as steady stimulation. Since each photic period is so long as to afford virtually fully asynchronous distribution of activity among channels, extensions of the null periods in the train (periods of no input between pulses) have no great effect upon the average brightness. Notice that a Brücke-Bartley effect is not indicated. This is because the PCF is not one allowing for it.

The lower left schema of Figure 5 also depicts what occurs when a smaller PCF of equal intensity is employed with a complete range of intermittencies. The chief points of difference are the lower peak of brightness when fused-pulse trains occupy a similar brief period. The Broca-Sulzer-type brightness fluctuation is less marked since the total flux acting in a summed fashion with PCF .33 is only one-half that for PCF .67. This means that fewer channels are being put into operation by the input. A more marked brightness deviation occurs below fusion. This is pictured as a peak considerably above that of steady illumination. Small PCFs combined with repetitions of 10 per second or lower provide for pulses short enough to preclude repetitive activation of channels and for null periods long enough to allow all activated channels to recover during the input cycle. In this respect, they are unlike PCF .67 since a short train cycled to produce a PCF .67 produces great synchrony in the channel of the optic pathway only at rates above fusion.

In using several PCFs and low repetitive rates, we are in another way producing the same conditions that gave rise to the Broca-Sulzer discovery. It is essentially immaterial that, in this case, flicker is a sensory end-result; whereas isolated flashes were observed by Broca and Sulzer. The phenomenon of Broca and Sulzer is simply one of the elements to be found in certain conditions for

flicker. The difference between Brücke-Bartley and Broca-Sulzer is not to be looked for at the level of cortical discharge pattern but only in terms of input timing. Broca and Sulzer used isolated pulses to produce synchrony and high amplitudes of cortical discharge, whereas the Brücke-Bartley effect requires a given relation between the null period in the stimulation cycle and pulse duration. Various writers have erroneously assumed the Brücke-Bartley effect to depend merely on pulse duration. Clearly, it does not depend sheerly upon this factor. Pulses of the same duration in trains having larger PCFs do not produce enhancement because the null period is too short to allow channels in operation to recover.

It follows also that PCFs very much smaller than any of those shown (i.e., a PCF of .01) presented in short trains above fusion would very likely not show much, if any, Broca-Sulzer-type brightening. On the other hand, below-fusion brightening would occur at far lower rates with PCF .01 than with PCF .33. At a rate of one cycle each three seconds, one would predict the brightness of the pulses to be higher than that of PCF .33 at 10 cps. By similar reasoning, one could predict approximately what intensity and stimulus rate above fusion would be required to match the brightness produced by a short stimulus train cycled so as to produce PCF .01 at one cycle each three seconds if the matching train were not affected by its predecessor and were distributed so as to give PCF .67. One might predict a match when the PCF-.67 trains were decreased by one-half and the entire train in question presented to the eye rapidly enough to reduce total train time to 33 msec.

#### G. SUMMARY

A number of findings and formulations exist in the literature on brightness produced either by continuous or intermittent inputs. They have never been put into a single description of optic-pathway activity. For this purpose, a mechanism that interrelates findings such as expressed in the Broca-Sulzer effect, the Brücke-Bartley effect, and the Talbot-Plateau law was described. Findings indicating that the Talbot-Plateau law fails for brief durations of intermittent input were briefly included.

#### REFERENCES

1. BALL, R. J. An investigation of chromatic brightness enhancement tendencies. Doctoral thesis, Michigan State University, East Lansing, 1963.
2. Bartley, S. H. Temporal and spatial summation of extrinsic impulses with the intrinsic activity of the cortex. *J. Cell. & Comp. Physiol.*, 1936, 8, 41-62.
3. ———. The neural determination of critical flicker frequency. *J. Exper. Psychol.*, 1937, 21, 678-686.
4. ———. A central mechanism in brightness discrimination. *Proc. Soc. Exper. Biol. & Med.*, 1938, 38, 535-536.

5. ———. Some aspects of intermittent photic stimulation. *J. Exper. Psychol.*, 1939, **25**, 462-480.
6. ———. Some factors in brightness discrimination. *Psychol. Rev.*, 1939, **46**, 337-358.
7. ———. The relation between cortical response to visual stimulation and changes in the alpha rhythm. *J. Exper. Psychol.*, 1940, **27**, 624-639.
8. ———. Vision, A Study of Its Basis. New York: Van Nostrand, 1941.
9. ———. Visual sensation and its dependence on the neurophysiology of the optic pathway. *Biol. Symp.*, 1942, **7**, 87-106.
10. ———. Light adaptation and brightness enhancement. *Percept. & Mot. Skills*, 1957, **7**, 85-92.
11. ———. A clarification of some of the procedures and concepts involved in dealing with the optic pathway. In Jung, R., & Kornhuber, H. (Eds.), *The Visual System: Neurophysiology and Psychophysics*. Heidelberg: Springer-Verlag, 1961.
12. BARTLEY, S. H., & NELSON, T. M. Some relations between sensory end-results and neural activity in the optic pathway. *J. of Psychol.*, 1963, **55**, 121-143.
13. BARTLEY, S. H., NELSON, T. M., & RANNEY, J. E. The sensory parallel of the reorganization period in the cortical response in intermittent retinal stimulation. *J. of Psychol.*, 1961, **52**, 137-147.
14. BARTLEY, S. H., NELSON, T. M., & SOULES, E. M. Visual acuity under conditions of intermittent illumination productive of paradoxical brightness. *J. of Psychol.*, 1963, **55**, 153-163.
15. BAUMGARDT, E. Sur les mécanismes de l'excitation visuelle. I. La théorie photochimique et la nature quantique et statistique de l'excitation lumineuse. *Arch. Sci. Physiol.*, 1947, **1**, 257-274.
16. BEAMS, J. W. The eye as an integrator of short light flashes. *J. Opt. Soc. Amer.*, 1934, **25**, 48.
17. BISHOP, G. H. Cyclic changes in the excitability of the optic pathway of the rabbit. *Amer. J. Physiol.*, 1933, **103**, 213-225.
18. BLONDEL, A., & REY, J. The perception of lights of short duration at their range limits. *Trans. Ill. Eng. Sec.*, 1912, **7**, 625.
19. BOYNTON, R. M. Some temporal factors in vision. In Rosenblith, W. A. (Ed.), *Sensory Communications*. New York: Wiley, 1961.
20. BROCA, A., & SULZER, D. La sensation lumineuse en fonction du temps pour les lumières colorées. *J. Physiol. Path. Gén.*, 1902, **4**, 632-640.
21. COHEN, J., & GORDON, D. A. Prevost-Fechner-Benham subjective colors. *Psychol. Bull.*, 1949, **46**, 97-136.
22. ECHLIN, F., & FESSARD, A. Synchronized impulse discharges from receptors in the deep tissues in response to a vibrating stimulus. *J. Physiol.*, 1938, **93**, 312-334.
23. GILMER, T. E. The integrating power of the eye for short flashes of light. *J. Opt. Soc. Amer.*, 1937, **27**, 386-388.
24. HELMHOLTZ, H. v. *Handbuch der Physiologischen Optik* (English trans., Ed., J. P. C. Southall). New York: Opt. Soc. Amer., 1924.
25. NELSON, T. M., & BARTLEY, S. H. The role of PCF in temporal manipulations of color. *J. of Psychol.*, 1961, **52**, 457-477.
26. NELSON, T. M., & BARTLEY, S. H. The Talbot-Plateau law and the brightness of restricted numbers of photic repetitions at CFF. *Vis. Res.*, 1964, in press.
27. NELSON, T. M., BARTLEY, S. H., & FORD, A. The brightness of part-spectrum targets stimulating the eye parafoveally at intermittency rates below fusion. *J. of Psychol.*, 1963, **55**, 387-396.
28. ———. Effects upon brightness produced by varying the length of the null

- interval separating successive "single" pulses: Sensory implications of the Alternation of Response Theory, I. *J. of Psychol.*, 1963, **56**, 99-106.
29. NELSON, T. M., BARTLEY, S. H., & JEWELL, R. M. Brightness changes associated with variation in the number of pulses in a repetitive photic train: Sensory implications of the Alternation of Response Theory, II. *J. of Psychol.*, 1963, **56**, 107-113.
  30. ———. Brightness from repetitive photic trains varying in pulse number and separation intervals: Sensory implications of the Alternation of Response Theory, III. *J. of Psychol.*, 1963, **56**, 175-183.
  31. ROSE, J. E., & MOUNTCASTLE, V. B. Touch and kinesthesia. In Magoun, H. N. (Ed.), *Handbook of Physiology (Sec. 1, Vol. 1)*. Washington, D.C.: *Amer. Physiol. Soc.*, 1959.
  32. STEWART, G. N. Is the law of Talbot true for very rapid intermittent light? *Proc. Roy. Soc. Edinb.* 1887-1888, **15**, 441-464.
  33. WEVER, R. C., & BRAY, G. W. Present possibilities for auditory theory. *Psychol. Rev.*, 1930, **37**, 365-385.

*Department of Psychology*  
*Michigan State University*  
*East Lansing, Michigan*



## DRUGS AND JUDGMENT: EFFECTS OF AMPHETAMINE AND SECOBARBITAL ON SELF-EVALUATION\*<sup>1</sup>

*Anaesthesia Laboratory of Harvard Medical School at Massachusetts  
General Hospital*

GENE MARSHALL SMITH AND HENRY K. BEECHER<sup>2</sup>

### A. INTRODUCTION

Various mood-altering drugs ("pep pills," "psychic energizers," "sedatives," "tranquilizers," etc.) are in widespread use. Many persons make decisions and judgments and engage in potentially dangerous activities, such as driving automobiles, after taking these drugs. The effects of such drugs on human decisions, judgments, and activities have merited study in the past and require further study (see references 1 to 18).

In the present investigation, judgment was studied by comparing a subject's performance on a 60-minute calculus test with his perception of that performance. Performance and subjective evaluation of performance occurred under drug and placebo conditions as described below. In earlier work (14), the method of comparing performance with perceived performance was found to be a highly sensitive indicator of judgment distortion produced in swimmers who had been given secobarbital sodium in a dose of 100 mg per 70 kg of body weight. The swimmers reported swimming faster after receiving secobarbital than after receiving placebo, but the *fact* was that secobarbital impaired performance to a highly significant degree.

### B. METHOD

#### 1. *Nature of the Self-Evaluations*

In each of five experimental sessions, after completing a 60-minute, 25-item calculus test, each subject estimated the number of problems he had solved correctly. As explained below, different forms of the test were used for the various sessions. Test scores were not reported to the subjects until the final experimental session was completed. In evaluating a given perform-

\* Received in the Editorial Office on June 25, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This work was supported in part by the United States Public Health Service, Contract No. USPH-M-987, in part by the Medical Research and Development Command of the United States Army, Contract No. DA-49-007-MD-2136, and in part by Mallinckrodt Chemical Works.

<sup>2</sup> The authors wish to acknowledge the helpful counsel of William G. Cochran, Professor of Statistics, Harvard University.

ance, a subject had recourse only to such ambiguous and subjective matters as how hard the problems seemed, how well he felt he had been functioning, how hard he had tried, how many answers he had guessed at, how many problems he had omitted, etc. A subject's judgment-error score for each experimental session was determined by subtracting the number of problems he had answered correctly from the number he estimated he had solved correctly. The subjects appeared challenged by the calculus tests, and their interest is indicated by the fact that all<sup>a</sup> asked to be informed of their scores at the end of the fifth experimental session. The problem of self-assessment also appeared to be approached by the subjects in a spirit of seriousness and cooperation.

## 2. Subjects

Seventy-eight male graduate and undergraduate students from Massachusetts Institute of Technology served as subjects. They ranged in age from 18 to 31, with a median of 21 years. Each subject was selected on the basis of his ability to reach a minimum criterion score on a test of intermediate calculus developed by the Educational Testing Service of Princeton, New Jersey.

## 3. Drugs and Drug Order

Each subject participated in five test sessions. He took amphetamine sulfate (14 mg per 70 kg of body weight) before two sessions, secobarbital sodium (50 mg per 70 kg of body weight) before one session, and placebo before two sessions. These dosages were selected on the basis of earlier work (14). All medications were oral. On each experimental day a subject took two capsules: one, two hours and 45 minutes before beginning the experimental session; the other, one hour and 15 minutes before beginning the session. On "amphetamine days," the first capsule contained amphetamine and the second contained a placebo. On secobarbital days, the first capsule contained a placebo and the second contained secobarbital. On placebo days both capsules contained placebos. The double-capsule technique was used because the peak effectiveness of secobarbital occurs earlier than does that of amphetamine. Subjects were asked to begin fasting one hour before taking the first capsule.

With respect to drug order, the subjects were divided into 12 groups with approximately seven subjects per group. Each group received medications in one of the following 12 orders: APSAP, APSA, ASPAP, ASPPA, PSAAP, PSAPA, PASAP, PASPA, SAPAP, SAPPA, SPAAP, SPAPA (A = amphetamine, P = placebo, S = secobarbital). In all orders, A and P occur

among the first three medications and occur again among the last two.<sup>3</sup> The first A and P are designated  $A_1$  and  $P_1$ ; the second A and P are designated  $A_2$  and  $P_2$ . Secobarbital occurs only among the first three medications in each order. The secobarbital results reported are based on comparisons of S and  $P_1$ . The drugs were distributed by a person not involved in collecting and evaluating the data; thus it was possible to maintain the double-blind procedure even though drug order was counterbalanced. Subjects were told that "pep pills" and placebos were being used. They were not told that a "sedative" was also being used.

#### 4. *Testing*

During each experimental session the testing was done on a group basis, in an auditorium, with vacant seats separating subjects. All testing sessions began at 5:45 p.m. Each evening the subjects took a different 60-minute calculus test. Six calculus tests<sup>4</sup> were prepared by Educational Testing Service. One, the screening test mentioned above, was designed to be slightly harder than the five<sup>1</sup> that were used for the experiment proper. Three of the remaining five were designed to parallel each other in difficulty and content; they were used for the first three sessions. The other two tests, which were used for the last two sessions, were designed to parallel each other and were similar to the first three, but were not necessarily parallel to the first three. Each of the six calculus tests contained 25 items and each item had five multiple-choice answers. The time limit was 60 minutes for each test, a time selected to be insufficient for completion of all items. All subjects received Form No. 1 of the calculus test in the first test session, Form No. 2 in the second test session, etc. Thus, test form and drug were balanced with respect to each other, while test form and session were not.

#### C. RESULTS

A subject's judgment-error score for each session was determined by subtracting the number of problems he had solved correctly from the number he estimated he had solved correctly. The placebo and amphetamine judgment-

<sup>3</sup> A and P were repeated a second time to permit an examination of the test-retest reproducibility of the amphetamine effects. Data concerning reproducibility (from several separate investigations) will be presented and discussed in a subsequent report.

<sup>4</sup> Some problems in each calculus test required knowledge of elementary algebra, trigonometry and analytic geometry. More advanced problems dealt with the differentiation and integration of elementary functions and with the application of these techniques to the computation of areas and volumes and to the solution of minimum and maximum problems. Some problems dealt with multiple integrals and the change of order of integration.

error scores of the 78 subjects are plotted as dots in Figure 1. The average of a subject's two placebo judgment-error scores  $[(P_1 + P_2) \div 2]$  is plotted on the ordinate, and the average of his two amphetamine judgment-error scores  $[(A_1 + A_2) \div 2]$  is plotted on the abscissa.<sup>5</sup> The horizontal line intercepting

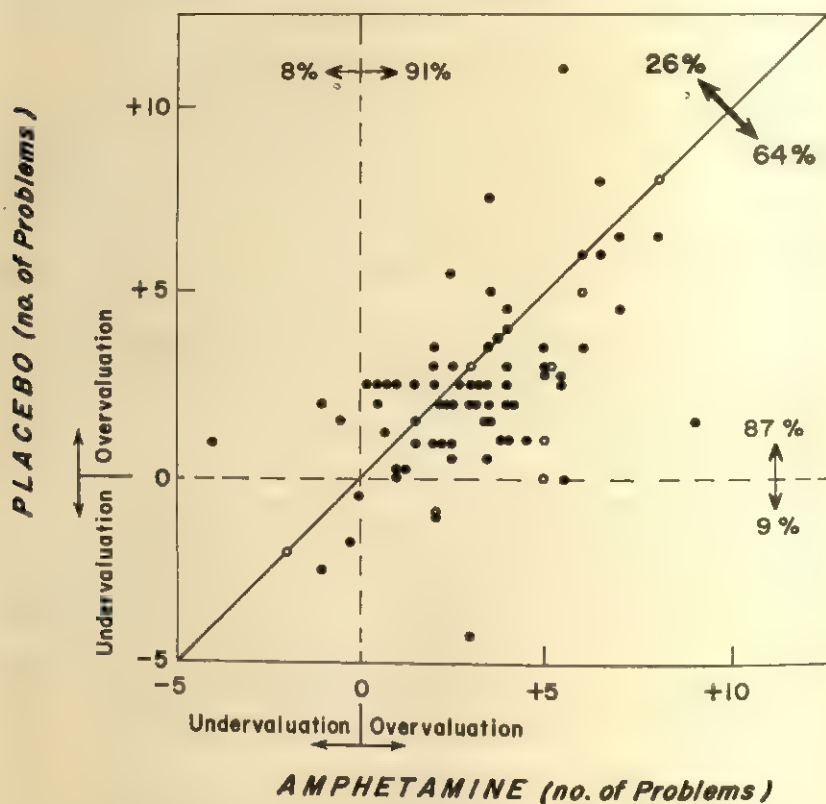


FIGURE 1

## JUDGMENT ERROR UNDER AMPHETAMINE AND PLACEBO CONDITIONS

Dots above the broken horizontal line indicate overvaluation with placebo. Dots to the right of the broken vertical line indicate overvaluation with amphetamine. Dots below the diagonal line indicate subjects who overvalued their calculus performance relatively more with amphetamine than with placebo. The eight open dots indicate scores based on a single amphetamine *vs.* placebo comparison rather than a mean of two such comparisons. See Footnote No. 5.

<sup>5</sup> Eight of the scores in the figure appear as open rather than solid dots. These are the scores of eight subjects who failed to give a performance estimate on one of the sessions. The position of each of these subjects in the figure is determined either by  $P_1$  and  $A_1$  or by  $P_2$  and  $A_2$ , rather than by  $[(P_1 + P_2) \div 2]$  and  $[(A_1 + A_2) \div 2]$ . The per cents reported in the figure, and the chi-square analyses reported in the text, are based on the entire group of 78 subjects represented in the figure.



zero on the ordinate separates the 68 subjects who overvalued their placebo performances from the seven who undervalued them. The vertical line intercepting zero on the abscissa separates the subjects who overvalued their amphetamine performances from those who undervalued them. As is shown in Figure 1, 87 per cent of the subjects overvalued their performance under the placebo condition and 91 per cent did so under the amphetamine condition.

Notwithstanding the highly significant tendency to overvalue placebo performances, a significant majority of the subjects had a *relatively greater* overvaluation score for their amphetamine performances than for their placebo performances. The 50 subjects below the diagonal line in Figure 1 overvalued their amphetamine performances relatively more than their placebo performances. The 20 subjects above that line overvalued their placebo performances relatively more. The deviation of 50/20 from 35/35 yields a

TABLE 1  
CALCULUS PERFORMANCE, ESTIMATE OF PERFORMANCE, AND JUDGMENT ERROR FOR  
FIVE EXPERIMENTAL SESSIONS: MEANS FOR 70\* SUBJECTS

Variable	A <sub>1</sub> **	P <sub>1</sub>	S	A <sub>2</sub>	P <sub>2</sub>
Calculus performance (Number of problems right)	14.10	14.43	13.51	13.81	13.97
Estimate of performance (Number of problems thought right)	17.44	17.04	16.72	16.57	16.16
Judgment Error*** (Number thought right minus number right)	3.34	2.61	3.21	2.76	2.19

\* Only the data of the 70 subjects having scores for all five sessions were used in computing the 15 means.

\*\* A<sub>1</sub> = first amphetamine session, P<sub>1</sub> = first placebo session, S = secobarbital session, A<sub>2</sub> = second amphetamine session, P<sub>2</sub> = second placebo session.

\*\*\* All five judgment-error means are significantly different from zero ( $p < 0.0001$ ).

chi-square value of 12.01 which, with a two-tailed test, is significant at the 0.001 level. Within the group of 58 subjects whose scores were in the direction of overvaluation for both the amphetamine and placebo conditions (and who did not fall on the diagonal line), 41 overvalued the amphetamine performances relatively more, and 17 overvalued the placebo performances relatively more. The deviation of 41/17 from 29/29 yields a chi-square of 9.12, which is significant at the 0.01 level. It is concluded that amphetamine produced an overvaluation tendency that is significantly greater than the highly significant placebo overvaluation tendency.

Table 1 reports the P<sub>1</sub>, A<sub>1</sub>, S, P<sub>2</sub> and A<sub>2</sub> means for each of three variables: number of problems solved correctly, number of problems the subjects thought



they solved correctly, and judgment error. Under the placebo condition the subjects, considered as a group, overestimated the number of correct solutions by about two and one-half problems. Under each of the two active drug conditions, the judgment error was approximately 30 per cent greater than that found with placebo.<sup>6</sup>

Table 2 shows the results of *t*-test evaluations of the effects of amphetamine and of secobarbital on judgment error. The  $[(A_1 - P_1) + (A_2 - P_2)] \div 2$  comparison was analyzed two ways: (a) For subjects missing either an  $A_1 - P_1$  or an  $A_2 - P_2$  score, the available difference score was used as an estimate of the average of the two difference scores. There are 78 subjects in that analysis. (b) Subjects with missing data in either the  $A_1 - P_1$  or the  $A_2 - P_2$  comparison were deleted. There are 70 subjects in that analysis. Both analyses yielded results consistent with those based on the chi-square

TABLE 2  
EFFECT OF AMPHETAMINE AND SECOBARBITAL, RELATIVE TO PLACEBO, ON JUDGMENT ERROR

Judgment error	<i>N</i>	$\bar{X}$	Statistic		
			<i>SD</i>	<i>t</i>	<i>p</i>
$[(A_1 - P_1) + (A_2 - P_2)] \div 2$	78	0.77	2.28	2.96	0.01
$[(A_1 - P_1) + (A_2 - P_2)] \div 2$	70	0.65	2.29	2.39	0.02
$S - P_1$	76	0.64	3.36	1.68	0.10

Legend:  $A_1$  = first amphetamine session,  $P_1$  = first placebo session,  $S$  = secobarbital session,  $A_2$  = second amphetamine session,  $P_2$  = second placebo session,  $N$  = sample size,  $\bar{X}$  = mean effect,  $SD$  = standard deviation,  $t$  = *t* value,  $p$  = two-tailed probability value.

analyses reported above. Although the  $S - P_1$  analysis did not reach the 0.05 level of significance, it should be mentioned that the magnitude of the secobarbital effect was as great as that found with amphetamine. Because amphetamine was given twice, its effect was assessed with greater precision (a smaller standard error) than was the effect of secobarbital. Thus, although the two effects were the same in magnitude, the amphetamine effect may be considered conclusive while the secobarbital effect must be considered only suggestive.

<sup>6</sup> The judgment error found during the later phase of the experiment ( $A_2 + P_2$ ) was significantly smaller ( $p < 0.05$ ) than that found earlier ( $A_1 + P_1$ ), but the meaning of this reduction is uncertain. As already stated, a subject was not told the number of problems he had solved correctly on any test until the fifth experimental session was concluded. Some subjects were observed working together on problems after each experimental session and this might have provided corrective feedback which contributed to the reduction in judgment error as the experiment progressed. Available data do not, however, permit a test of that possibility.

## D. DISCUSSION

While it is of interest that a large per cent of subjects (87 per cent) overvalued their calculus performance under the placebo condition, our main concern in the present discussion is with the effect of amphetamine on judgment and with the implications of that effect. A subsequent report will concern the overvaluation tendency found under placebo and nondrug conditions in this and other studies.

The fact that the effect of amphetamine on judgment was in the direction of impairment rather than improvement is of interest for three reasons.

First, considerable evidence is available demonstrating improvement in performance on various mental and physical tasks after administration of amphetamine, but its potential impairment effects are not well documented. In concluding their review of 112 publications concerning effects of caffeine and the amphetamines on human performance, Weiss and Laties (18, p. 32) say:

Except for reports of insomnia, the subjective effects of the amphetamines in normal doses are usually favorable. Moreover, no one has ever presented convincing evidence that they impair judgment.

Second, amphetamines are widely used by ambulant patients and by individuals who take them without prescription; hence indication of any type of judgment impairment produced by such agents has important practical implications. The potential significance of this matter in regard to juvenile delinquency and crime has been discussed elsewhere (13).

Finally, the impairment of judgment shown above is not paralleled by the effects of amphetamine on the performance of the same group of 78 subjects tested with respect to two other aspects of mental functioning. In an earlier report, amphetamine was shown to facilitate the performance of these subjects on a clerical test of digit-letter coding to a highly significant degree ( $p < 0.0001$ ) and to have no effect on their calculus performance (15). Thus, when we examine the effects of amphetamine on the same group of 78 subjects evaluated with regard to three different types of mental performance (judgment, coding, and calculus), we find one type of performance significantly impaired, one type significantly improved, and one type not affected. These results stress the need for cautious generalization concerning drug-induced changes in mental performance. §

Our analyses suggest that secobarbital might also be capable of elevating the overvaluation tendency beyond that found under the placebo condition. That a small dose of secobarbital (50 mg is one-half the standard sedative dose) might produce certain mental effects similar to those produced by amphetamine

(a "stimulant") is of interest and is consistent with earlier findings. Smith and Beecher (13) administered amphetamine and secobarbital to athletes in the same dosages as those employed in the present study and found that both agents produced significant increases in reports of mental and physical activation, elation, and intoxication. The activation effects, however, were more definite in the case of amphetamine than in the case of secobarbital.

The nature of the drug-induced judgment error found in the amphetamine results (and suggested in the secobarbital results) is not well understood at present. Both agents produce significant mood changes which might play a role in this matter. In addition to the activation, elation, and intoxication effects just mentioned, amphetamine (in the 14 mg per 70 kg body-weight dosage) has been found to produce significant increases in reports of boldness, friendliness, and impulsiveness, both in athletes (13) and in the 78 subjects of the current study (15). Further work is needed to elaborate the meaning and significance of the observed association between drug-induced changes in mood and judgment.

### E. SUMMARY

Each of 78 graduate and undergraduate college students attempted to solve 25 calculus problems and afterward estimated the number of his correct solutions. Subjects did this on each of five occasions: twice after taking amphetamine sulfate (14 mg per 70 kg of body weight), once after taking secobarbital sodium (50 mg per 70 kg of body weight), and twice after taking a placebo. After taking placebo the subjects overestimated, to a significant degree, the number of problems solved correctly. After taking amphetamine or secobarbital the overestimation tendency, or judgment error, was still greater. The increase in judgment error was statistically significant with amphetamine but only suggestive with secobarbital. Implications of the findings concerning the effect of amphetamine on judgment were discussed.

### REFERENCES

1. GOODNOW, R. E., BEECHER, H. K., BRAZIER, M. A. B., MOSTELLER, F. C., & TAGIURI, R. Physiological performance following a hypnotic dose of a barbiturate. *J. Pharmacol.*, 1951, **102**, 55-61.
2. HAUTY, G. T., & PAYNE, R. B. Effects of dextro-amphetamine upon judgment. *J. Pharmacol.*, 1957, **120**, 33-37.
3. KELLY, E. L., MILLER, J. G., MARQUIS, D. G., GERARD, R. W., & UHR, L. Continued meprobamate and prochlorperazine administration and behavior. *A.M.A. Arch. Neurol. & Psychiat.*, 1958, **80**, 247-252.
4. KORNETSKY, C. Effects of meprobamate, phenobarbital and dextro-amphetamine on reaction time and learning in man. *J. Pharmacol.*, 1958, **123**, 216-219.
5. KORNETSKY, C., & HUMPHRIES, O. Psychological effects of centrally acting drugs

- in man: Effects of chlorpromazine and secobarbital on visual and motor behavior. *J. Ment. Sci.*, 1958, **104**, 1093-1099.
6. KORNETSKY, C., HUMPHRIES, O., & EVARTS, E. V. Comparison of psychological effects of certain centrally acting drugs in man. *A.M.A. Arch. Neurol & Psychiat.*, 1957, **77**, 318-324.
  7. LEHMANN, H. E., & CSANK, J. Differential screening of phrenotropic agents in man: Psychophysiological test data. *J. Clin. & Exper. Psychopath.*, 1957, **18**, 222-235.
  8. LOOMIS, T. A., & WEST, T. C. Comparative sedative effects of barbiturate and some tranquilizer drugs on normal subjects. *J. Pharmacol.*, 1958, **122**, 525-531.
  9. MARQUIS, D. G., KELLY, E. L., MILLER, J. G., GERARD, R. W., & RAPOPORT, A. Experimental studies of behavioral effects of meprobamate on normal subjects. *Ann. N. Y. Acad. Sci.*, 1957, **67**, 701-711.
  10. REITAN, R. M. Comparative effects of placebo, ultran, meprobamate on psychologic test performances. *Antibiot. Med.*, 1957, **4**, 158-165.
  11. SEASHORE, R. H., & IVY, A. C. Effects of analeptic drugs in relieving fatigue. *Psychol. Monog.*, 1953, **67**, 1-16.
  12. SMITH, G. M., & BEECHER, H. K. Amphetamine sulfate and athletic performance: I. Objective effects. *J.A.M.A.*, 1959, **170**, 542-557.
  13. ———. Amphetamine, secobarbital, and athletic performance: II Subjective evaluations of performances, mood states, and physical states. *J.A.M.A.*, 1960, **172**, 1502-1514.
  14. ———. Amphetamine, secobarbital, and athletic performance: III. Quantitative effects on judgment. *J.A.M.A.*, 1960, **172**, 1623-1629.
  15. SMITH, G. M., WEITZNER, M., LEVENSON, S. R., & BEECHER, H. K. Effects of amphetamine and secobarbital on coding and mathematical performance. *J. Pharmacol.*, 1963, **141**, 100-104.
  16. STEINBERG, H. Selective effects of an anaesthetic drug on cognitive behaviour. *Quart. J. Exper. Psychol.*, 1954, **6**, 170-180.
  17. VON FELSINGER, J. M., LASAGNA, L., & BEECHER, H. K. The persistence of mental impairment following a hypnotic dose of a barbiturate. *J. Pharmacol.*, 1953, **109**, 284-291.
  18. WEISS, B., & LATIES, V. G. Enhancement of human performance by caffeine and the amphetamines. *Pharmacol. Rev.*, 1962, **14**, 1-36.

*Department of Anaesthesia*  
*Massachusetts General Hospital*  
*Boston 14, Massachusetts*



## BEHAVIOR IN THE ATTAINMENT OF ECONOMIC CONCEPTS: II.\*

*College of Education, University of Georgia*

---

RACHEL S. SUTTON

---

### A. INTRODUCTION

A concept is introduced into language when certain kinds of phenomena are selected for attention. There is neither a set of rules nor a timetable for learning concepts. The particular point at which a concept becomes clear to a child cannot be set in advance by the curriculum. In any subject area, the learner needs a firm and broad grasp of established knowledge and techniques. These he can be taught. Imagination, ingenuity, and discovery he can develop. The task of the teacher becomes that of guiding the learner in the attainment of significant concepts.

Concept attainment is an important step in structuring knowledge. Concepts fill out the framework of a discipline. Breadth and depth are necessary in undetermined amounts. The arrangements of man's experience are constantly put to test by exigencies of time and demand. Man's past experience is a reservoir to be drawn upon continuously in building concepts.

The process of forming concepts is important, as well as the concepts themselves. Whatever calls forth a human response and acts as fermentation for subsequent responses may be regarded as material for concept formation. It is valuable also to the individual to know that he has an active part in the structuring of knowledge. To fulfill this function, he must retain the attitude of the adventurer and explorer; develop careful, precise workmanship; and examine, modify, and add to concepts that categorize the vast store of human experience.

The compatibility of concepts determines the level of relationships. For example, the concepts of freedom and slavery cannot exist side by side without destructive abrasions and confusion. An individual makes his own contribution to a structure of knowledge by detecting the incompatibility and absurdities in existing concepts and by discovering newer ways of attaining compatibility and convergence.

Economics is an area in which concepts can be constantly observed and

---

\* Received in the Editorial Office on June 25, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.



tested. It offers an opportunity for exploration and discovery of attributes that place an experience or object in one category or another. It has no history of being taught as a subject in the elementary school nor of being presented logically. The elementary-school child learns incidentally about the powerful economic forces that affect all of human life.

### B. RELATED RESEARCH

Concepts and the language\* that infuses and instruments them give power to cognitive activity (11). The capacity to impose superordinate structures in the interest of observing matter move simply and deeply is a powerful tool of human intelligence. The higher meanings transform the lower, concrete ones. The child does not have to restructure separately all of his earlier meanings. Once a new structure has been incorporated into his thinking, it gradually spreads to the older meanings as they are drawn into the intellectual operations of the abstract type. The child has the capacity to understand higher-order structures that replace and give new power to the conceptual structures that were used in the beginning toward a mastery of abstractions.

In a study reported by Sutton (7), children were at an initial stage of development in their understanding of money and the accumulation of capital. It was assumed that capital accumulation in the American economy takes place by a two-sided process of saving money, or of withholding current money income from expenditure on current consumption and of investing the current money income in a stock of valuable goods to produce more capital.

To subjects in the investigation, money was a thing to spend for the near-at-hand and the immediate. A bank was a safe place to keep money until it was needed for spending. Older children were slightly less egocentric than younger ones. The majority of replies were on the precategorical level. Age, intelligence, and socioeconomic background had little effect on the understanding of the production credit of money. Older, more intelligent children tended to moralize and explain in terms of rightness and wrongness. There was little evidence that the respondents had been stimulated by their environment to explore the creative uses of capital and the numerous ways by which it could be acquired. Six stages of growth in the formation of economic concepts suggested levels of thought, from no cognitive response to the significance of a single act, in terms of its position in a system of relationships.

The stages of conceptual development do not owe their appearance simply to internal maturation, but depend to a large measure on the intellectual materials with which the child is provided. Mere provision of materials is not enough. An active educational process is required to support the integration

of cognitive experience (8). Connectives of similarities and differences, time sequence, cause-effect relationships, evidence, and judgment help a child deal with an ordered world in terms of logic and a perception of reality.

The more mature and intelligent the human individual (1), the more abstract are his symbolic processes, as exemplified by the hosts of fine discriminations that characterize his language behavior. The place of the organism in the evolutionary scale is indicated by the quality of discriminations the organism can usually make. The greater the discriminatory capacity of the organism, the more involved are detachable reactions finally included in his meditation process.

An opportunity for studying attainment of significant concepts is found in the relatively unexplored areas of the quality of children's responses to words. Due to acquisition of greater understanding (2), the character of definitions of words change with increased maturity. Recall and recognition techniques for the measurement of precise knowledge of word meanings have a limited number of common factors (3, 4). Russell (5, 6) suggested a developmental sequence in concept attainment as sensorimotor, concrete-perceptual, functional, and abstract-conceptual.

### C. PROCEDURE

Qualitative studies of the kinds of meanings children acquire and use at various ages in specific subjects are needed. The present study is concerned with these questions: (a) When a child of a given age is confronted with several alternative meanings of words in a relatively unexplored area, such as economics, which level of meaning—incorrect, concrete, functional, or abstract—does he choose as the best meaning? (b) After instruction in the area, is there a difference in his choice of meaning?

One hundred words embodying economic thought were chosen from materials prepared for children—books, newspapers, and magazines. They were checked with *The Teacher's Word Book of 30,000 Words* (10). Each word was ascribed four meanings illustrative of different levels of cognition.

A pilot test was constructed and administered to 12 persons with an academic background in economics. They were instructed to classify the meanings of each term as (a) incorrect, (b) concrete, (c) functional, and (d) abstract. From the results of these ratings, revisions were made and copies given to six distinguished economists who agreed to serve as validating jurors.

The test was then arranged for four levels as determined by grades: 10 items for the first grade, 20 for the third, 60 for the fifth, and 100 for adult populations (9).

The selected subjects were children in the first, third, and fifth grades of an elementary school in Oglethorpe County, Georgia, and elementary-school teachers with a limited background in economics, who were studying for fifth- and sixth-year certificates during the 1963 summer session at the University of Georgia. Each subject was given a copy of the test. The investigator read the items to the children in each grade.

The children were instructed in economics over a period of six months. A research assistant with a background in economics and experience in the elementary school met each grade two afternoons a week for 30 minutes. Materials prepared by the Joint Council on Economic Education, newspapers, magazines, films, and filmstrips were used to answer children's questions and to satisfy their curiosities in this area. The same test was used after instruction.

#### D. RESULTS

Data treated by an analysis of variance with angular or inverse sine ( $\arcsin$ ) transformation and summarized in Tables 1 and 2 accepted the hypotheses:

TABLE 1  
EFFECT OF AGE ON CONCEPT ATTAINMENT

Category of response	Source of variation	<i>df</i>	Sum of squares	Mean squares	<i>F</i>
Incorrect	Among levels	4	5145.42	1286.35	3.10*
	Within levels	105	43542.25	414.68	
	Total	109	18161.90		
Concrete	Among levels	4	1483.95	370.99	3.62*
	Within levels	105	10757.39	102.45	
	Total	109	12241.24		
Functional	Among levels	4	1041.01	260.25	1.35
	Within levels	105	20240.80	192.77	
	Total	109	10524.36		
Abstract	Among levels	4	6158.12	1539.53	5.29**
	Within levels	105	30525.75	290.72	
	Total	109	16442.97		

\* Significant at the .10 level.

\*\* Significant at the .05 level.

(a) that incorrect and concrete responses decrease with age, and abstract responses increased; and (b) that incorrect and concrete responses decrease significantly after planned instruction, and functional and abstract responses increased.

#### E. SUMMARY

Earlier investigations pointed up the need to assess the effect of age and instruction on concept attainment. The present study assumed that kinds of

TABLE 2  
EFFECT OF INSTRUCTION ON CONCEPT ATTAINMENT

Category of response	Source of variation	df	Sum of squares	Mean squares	F
Incorrect	Grade	2	1924.75	926.38	14.60**
	Pre- <i>vs.</i> -post in grade	3	5021.77	1673.92	
	Error	118	13531.11	114.67	
	Total	123	20477.63		
Concrete	Grade	2	999.07	499.53	19.71**
	Pre- <i>vs.</i> -post in grade	3	7786.76	2595.58	
	Error	118	15543.48	131.72	
	Total	123	24329.29		
Functional	Grade	2	639.81	319.91	17.08**
	Pre- <i>vs.</i> -post in grade	3	3589.08	1196.36	
	Error	118	8264.11	70.03	
	Total	123	12492.87		
Abstract	Grade	2	2971.73	1485.86	16.20**
	Pre- <i>vs.</i> -post in grade	3	3761.28	1253.76	
	Error	118	9132.78	77.40	
	Total	123	15865.63		

\*\* Significant at the .05 level.

word meanings—incorrect, concrete, functional, and abstract—reflected levels of thought and stages in concept development. Results indicate (a) that incorrect and concrete responses decrease with age, and abstract responses increase; and (b) that incorrect and concrete responses decrease after planned instruction, and functional and abstract responses increase.

#### REFERENCES

1. FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. Psychol. Monog.*, 1949, **39**, 151-204.
2. FEIFEL, H., & LORGE, I. Qualitative differences in the vocabulary responses of children. *J. Educ. Psychol.*, 1950, **41**, 1-18.
3. HURLBURT, D. The relative values of recall and recognition techniques for measuring precise knowledge of word meanings; Nouns, verbs, and adjectives. *J. Educ. Res.*, 1954, **47**, 561-576.
4. KRUGLOV, L. Qualitative differences in the vocabulary choices of children as revealed in a multiple-choice test. *J. Educ. Psychol.*, 1953, **44**, 229-243.
5. RUSSELL, D. Dimensions of children's meaning vocabularies in grades four through twelve. *Univ. Calif. Publ. Educ.*, 1954, **11**, 315-414.
6. RUSSELL, D., & SAADEH, I. Qualitative level in children's vocabularies. *J. Educ. Psychol.*, 1962, **53**, 170-174.
7. SUTTON, R. Behavior in the attainment of economic concepts. *J. of Psychol.*, 1962, **53**, 37-46.
8. ———. Words *vs.* concepts. *Education*, 1963, **83**, 537-540.
9. ———. Attainment of economic concepts. Unpublished paper read at AERA, Chicago, February 20, 1964.
10. THORNDIKE, E. L., & LORGE, I. The Teacher's Word Book of 30,000 Words. New York: Bur. Publications, Teachers Coll., Columbia Univ., 1959.

11. VYGOTSKY, L. *Thought and Language* (ed. & trans. by E. Hanfmann & G. Vaker).  
Cambridge, Mass.: M. I. T. Press, 1962.

*College of Education*  
*University of Georgia*  
*Athens, Georgia*



## STEPWISE DISCRIMINANT-FUNCTION ANALYSIS OF SEMINARY-CANDIDATE MMPI SCORES\*<sup>1</sup>

*Department of Psychology, Boston College*

---

DANIEL J. BAER AND JAMES F. MOYNIHAN, S.J.

---

### A. PURPOSE

Although the MMPI has been extensively used for the diagnosis of severe psychopathology, it has been less successful as a screening device when applied to a grossly normal population. Among its critics, several investigators have questioned the value of the test (other than as a gross screening measure) in the selection of seminary candidates (4, 10, 11).

However, a more detailed examination of test performance within the seminary group could be useful (8). Each year several students leave the seminary because of poor psychological adjustment to their environment, even though MMPI scores have revealed no apparent pathology. Because differentiation between those who remain and those who leave the seminary is essentially a classification problem, it seems possible that a discriminant-function analysis of MMPI scores might reveal group differences so far not observed by the univariate *t* tests reported in the literature (4, 5, 10). The purpose of the present study was to determine, by means of stepwise linear discriminant-function analysis (LDFA), whether those students who leave and those who remain in the seminary can be significantly differentiated in their performance on 10 MMPI scales.

### B. METHOD

#### 1. *Subjects*

The validation sample consisted of 202 candidates who were accepted into the New England Province of the Society of Jesus during the years 1956 to 1961. All of the Ss were white, American-born, college-age students and, at the time of the LDFA, 140 Ss (Group 1) were still members of the Society while 62 Ss (Group 2) had left. A cross-validation sample consisted of 64 candidates who were accepted during the years 1962 and 1963. At the time

---

\* Received in the Editorial Office on July 7, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> We wish to thank Rev. W. A. Barry, S.J., for assistance in the data processing.

of the analysis 48 Ss (Group 3) were still members of the Society while 16 Ss (Group 4) had left the seminary.

## 2. Test Instrument

The test scores were based on the nine clinical scales and the K validity scale of the Bier modified version of the MMPI (1, 2). This version consists of 366 items of the original 550-item MMPI; however, 37 items of the first 366 items of the original MMPI have been replaced by 37 items selected from the remaining pool of 184 items. This modification was made to insure a greater discrimination among seminary-student applicants.

## 3. Procedure

A stepwise LDFA was applied to the 10 scales of the MMPI for the 140 Ss of Group 1 and the 62 Ss of Group 2. The derived lambda weights were then used in a linear discriminant equation to classify the 48 Ss of Group 3 and the 16 Ss of Group 4. A chi-square test of the independence of the observed number *vs.* the predicted number of Ss falling into Groups 3 and 4 was used as an estimate of the success of the LDFA as a classification device for seminary-student applicants.

## C. RESULTS

The Ss' MMPI scores were punched on IBM cards, and the data were processed on an IBM 1620 computer. Table 1 shows the means, *SD*s and *t* tests of the differences between the means of Groups 1 and 2 for the 10 scales of the MMPI. It may be seen that only the K- and Sc-scale means are significantly different (at the .05 level) in the two groups.

TABLE 1  
MEANS, *SD*s, AND *t* TESTS OF DIFFERENCES BETWEEN MEANS FOR 10 MMPI SCALES

Scale	Group 1		Group 2		<i>t</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>	
K (test-taking attitude)	17.09	4.24	18.31	3.77	1.97*
Ms (hypochondriasis)	2.73	2.57	2.18	2.14	1.47
D (depression)	17.09	3.81	16.81	2.99	0.52
Hy (hysteria)	20.44	4.02	21.15	3.76	1.16
Pd (psychopathic deviate)	16.76	3.12	17.16	2.99	0.84
Mf (masculinity and femininity)	29.71	5.06	29.10	5.38	0.71
Pa (paranoia)	9.40	2.44	9.66	1.93	0.74
Pt (psychasthenia)	7.65	4.95	6.59	4.57	1.42
Sc (schizophrenia)	10.07	4.59	8.24	3.78	2.74**
Ma (hypomania)	15.54	3.76	15.65	3.78	0.91

\*  $p < .05$ .

\*\*  $p < .01$ .

The data were then submitted to a stepwise LDFA (6). Table 2 shows the multiple point-biserial  $R$ ,  $df$ , and  $F$  ratio for each step in the analysis

TABLE 2  
STEPWISE DISCRIMINANT-FUNCTION ANALYSIS  
BASED ON 10 MMPI SCALES

Step	$R$	$df$	$F$	Scale eliminated
1	.403	10,191	1.28	D (depression)
2	.403	9,192	1.43	K (test-taking attitude)
3	.403	8,193	1.63	Pd (psychopathic deviate)
4	.375	7,194	1.81	Pa (paranoia)
5	.348	6,195	2.03	Mf (masculinity and femininity)
6	.328	5,196	2.38*	Pt (psychasthenia)
7	.278	4,197	2.73*	Ma (hypomania)
8	.226	3,198	3.29*	Hs (hypochondriasis)
9	.176	2,199	4.36*	Hy (hysteria)
10	.129	1,200	7.48**	

\*  $p < .05$ .

\*\*  $p < .01$ .

and the variable that was eliminated for the given step. Thus for Step 1 the obtained  $F$  of 1.28 is not significant beyond the .05 level, and the multiple  $R$  of .403 is a descriptive measure of the degree to which it is possible to predict attrition-survival. The insignificant  $F$  indicates that the two groups did not differ in their MMPI profiles based on the 10 scale scores. The variable that contributed least to this discrimination (D) was then eliminated in Step 2. This same procedure was followed, eliminating one scale at a time from the remaining number of scales, until all of the scales but one had been selected. It may be seen that, by Step 6 (and in subsequent steps), significant discrimination occurred in the MMPI profiles of the two groups when each scale was removed from the battery. Table 3 shows the lambda coefficients derived from Steps 6 through 9. These weights were then applied in an LDFA equation<sup>a</sup> in order to classify the  $S$ s of the cross-validation sample.

TABLE 3  
LAMBDA COEFFICIENTS AND LINEAR DISCRIMINANT-FUNCTION EQUATIONS  
FOR STEPS 6 TO 9

Step	Midpoint	Linear discriminant-function equations
6	255.31	— 23.30 Sc + 15.43 Hy — 23.81 Hs + 8.78 Ma + 9.73 Pt
7	252.50	— 16.56 Sc + 14.93 Hy — 19.63 Hs + 9.07 Ma
8	59.66	— 15.23 Sc + 14.04 Hy — 17.78 Hs
9	8.24	— 19.52 Sc + 8.99 Hy

The midpoint score was determined by summing the products of the lambda weights for each scale by the corresponding group means for the same scale. The value halfway between the sum for Group 1 and Group 2 represented the decision point for classifying the *Ss*. When an *S's* score fell above this point, it was predicted he would fall into Group 3 (i.e., be a survivor), while if his score fell below this value it was predicted he would fall into Group 4 (i.e., be a terminator). Since multiplying each *S's* score by just one lambda weight would not change the rank order of the *Ss* from their raw-score performance, and since only one lambda weight was associated with Step 10, its value was not necessary for this step and is not reported in Table 3.

Table 4 shows the chi squares based on the tests of independence of the predicted and observed number of *Ss* falling into Groups 3 and 4. Since none of the chi squares is significant, it seems that there is no relationship between predicted and observed classification of *Ss*.

TABLE 4  
OBSERVED AND PREDICTED NUMBER OF *Ss* CLASSIFIED INTO GROUPS 3 AND 4 AND  
CHI-SQUARE TESTS OF INDEPENDENCE OF TWO MODES OF CLASSIFICATION

Step	Observed classification	Predicted classification		Chi square
		Group 3	Group 4	
6	Group 4	30	18	3.48
	Group 3	14	2	
7	Group 4	30	18	1.66
	Group 3	13	3	
8	Group 4	28	20	1.80
	Group 3	12	4	
9	Group 4	30	18	1.66
	Group 3	13	3	
10	Group 4	28	20	0.56
	Group 3	11	5	

#### D. DISCUSSION

On the basis of the significant *F* ratios reported in Table 2, the LDFA appeared to be a promising method of predicting the attrition and survival of seminary students. However, when the lambda weights were applied to the cross-validation sample, the number of *Ss* successfully classified was not beyond chance expectancy. Thus LDFA was not a consistent discriminator within the seminary-student groups.

It seems that this failure is in part due to the inherent limitation of the MMPI clinical scales when used with a grossly normal population. Perhaps with new scales, developed from an item analysis of MMPI responses of *Ss*

who have left and Ss who have remained in the seminary, a more successful classification would be possible. Some progress along this line (1) has already been accomplished. A supplementary if not alternate approach would be to establish a series of decision rules, such as those recommended by Kleinmuntz (7), using the clinical scales of the MMPI for the classification of Ss. However, because of the inherent limitation associated with any paper-and-pencil measure of personality, it would seem that investigators should also be encouraged to develop new objective tests rather than relying on the convenient and popular (yet time consuming and so far nondiscriminating) MMPI.

A second obvious yet not easily resolved problem that may have led to the present results is relevant to the criteria for "success" in the seminary. Although perseverance in the seminary may be the simplest and most objective criterion, it may not be the one best suited for this type of personal-selection problem. Vaughan (10) indicates that, although religious supervisors tend to attribute most of the "failures" to some kind of psychological problem, some of those who leave may have better mental health than those who remain. In either case, the assumptions that those who remain are in some ways different psychologically from those who leave the seminary and that "adjustment" is a pervasive and unitary trait accounting for perseverance have never been well established (4, 5). Perhaps a nonlinear relationship exists between the criterion of perseverance in the seminary and psychological adjustment. Thus, an optimal range of adjustment would be necessary for success in the religious life, with individuals having more or less of this quality being unsuited (or at some disadvantage).

A third unavoidable difficulty associated with the outcome of the present analysis was the tendency to misclassify Ss in an LDFA when the number of individuals falling into each group is greatly different (3). In this case, Group 1 (with an  $N$  of 140) was over twice the size of Group 2 (with an  $N$  of 62). Such a result enhanced the likelihood of Ss being overassigned to the smaller group. Thus, more Ss were predicted to leave than did leave the seminary simply because of the original disparity in group sizes. However, since the cross-validation sample consisted only of candidates who had completed their first one or two years of study, it is possible that a few more of the seminary students who were predicted to leave will eventually do so, thus increasing the number of correct classifications.

From a comparison of the data in Tables 1 and 2, it may be seen that, whereas the  $K$  scale significantly differentiated Groups 1 and 2 (using a  $t$  test of the difference between means), the same scale was the second least



discriminating in an LDFA. Such a result is evidence that, in LDFA, the better discriminators in a test battery need not be the better individual discriminators. Also, from Table 3 it may be seen that the magnitude of the lambda weight associated with each scale (i.e., the relative importance of each scale) changed with each step in the LDFA. Thus, in stepwise LDFA, some consideration should be given to the question of which set of lambda weights should be used in classifying Ss. Rather than include the step with the maximum number of significant variables, it may be more practical for the investigator to select a step that has a fewer number of scales yet is just as successful in classification.

### E. SUMMARY

Test scores for 140 students who remained and 62 students who left the seminary were subjected to a stepwise linear discriminant-function analysis to determine if the two groups differed in their MMPI profiles. Although by Step 6 five scales—Pt (psychasthenia), Ma (hypomania), Hs (hypochondriasis), Hy (hysteria), and Sc (schizophrenia)—significantly differentiated the two groups (at the .05 level), when the derived lambda coefficients were applied in a linear discriminant-function equation with test scores for a cross-validation sample of 64 Ss, the number of correctly classified Ss was not significantly greater than chance expectancy. The authors conclude that the nine clinical scales and the K validity scale of the MMPI cannot distinguish those students who leave the seminary from those who remain.

### REFERENCES

1. BARRY, W. A. A seminary scale for Bier's modified MMPI. Unpublished Master's thesis, Fordham University, New York, 1960.
2. BIER, W. C. A comparative study of five catholic college groups on the MMPI. In G. S. Welsh & W. D. Dahlstrom (Eds.), *Basic Readings on the MMPI in Psychology and Medicine*. Minneapolis: Univ. Minn. Press, 1956. Pp. 586-609.
3. COOLEY, W. C., & LOHNES, P. R. *Multivariate Procedures for the Behavioral Sciences*. New York: Wiley, 1962.
4. FEHR, R. C. An inventory and a projective personality study of a religious and a lay group. Unpublished Doctoral dissertation, Fordham University, New York, 1960.
5. FRIEDL, F. P. Vocational interests of successful and unsuccessful seminarians in a foreign-mission society. Unpublished Master's thesis, Catholic University, Washington, D.C., 1952.
6. HURST, R. L., & WISER, G. Discriminant function analysis (Stepwise). IBM 1620 program 6.0.076. New York: Internat. Bus. Machine Corp., 1962.
7. KLEINMUNTZ, B. MMPI decision rules for the identification of college maladjustment. *Psychol. Monog.*, 1963, *77*, Whole No. 577.
8. MCCARTHY, T. N., & DONDERO, E. A. Predictive variables and criteria of success in religious life: Needed research. *Catholic Psychological Rec.*, 1963, *1*, 71-80.

9. MURRAY, J. B. Training for the priesthood and personality and interest test manifestations. Unpublished Doctoral dissertation, Fordham University, New York, 1957.
10. VAUGHAN, R. P. A psychological assessment program for candidates to the religious life: Validation study. *Catholic Psychological Rec.*, 1963, 1, 65-70.
11. WAUCK, L. A. An investigation into the use of psychological tests as an aid in the selection of candidates for the diocesan priesthood. Unpublished Doctoral dissertation, Loyola University, Chicago, 1956.
12. WEISGERBER, C. A. Survey of a psychological screening program in a clerical order. In M. B. Arnold, P. Hispanicus, C. A. Weisgerber, & P. F. D'Arcy (Eds.), *Screening Candidates for the Priesthood and Religious Life*. Chicago: Loyola Univ. Press, 1963. Pp. 107-148.

*Department of Psychology*  
*Boston College*  
*Chestnut Hill 67, Mass.*

## HUMAN-FIGURE DRAWINGS IN RELATION TO MODERNIZATION IN SUDAN\*<sup>1</sup>

*American University of Beirut and Brooklyn College*

---

MALIK B. BADRI AND WAYNE DENNIS

---

### A. INTRODUCTION

From a consideration of drawings of a man made by children in a heterogeneous assortment of groups, it appears probable that children in nearly all groups draw the kinds of men of which they approve. While a brief report has appeared (1), most of the data supporting this position are as yet unpublished. If this principle is upheld, human-figure drawings, and drawings of other subjects, may provide easily obtained and indirect indices of social attitudes and of changes in attitudes.

One of the topics that may be elucidated by this method is the nature of some of the psychological changes that occur in the course of the modernization of a nation or of a smaller social unit.

The term "modernization" refers to a complex of changes that take place as a group progresses from a more primitive, traditional way of life to one based largely on modern industry and technology. The visible signs of modernization include products such as locomotives, automobiles, airplanes, paved streets and highways, refrigerators, plumbing, telephones, cinemas, and television. Accompanying the introduction of these artifacts are many other changes, such as an increase in literacy and education, increases in health facilities, new forms of recreation and of employment, changes in child-rearing practices, etc. Many of these social changes take place concurrently, so that a community which is advanced in one respect, is likely also to be advanced in other aspects of modern life.

It may be assumed that the changes which constitute modernization do not take place voluntarily unless positive attitudes toward modernization are present in the population which is undergoing change. It is in the measurement of these attitudes that we are interested.

In this context, drawings of the human figure provide a useful research tool. In a traditional society, the approved-of-man often has an appearance

---

\* Received in the Editorial Office on July 7, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Expenses in this study were defrayed by a grant from the Rockefeller Brothers Fund to the American University of Beirut.

that distinguishes him from the men of other groups. The distinctive items of appearance may involve hairstyle and headgear, the treatment of facial hair, scarification, and, of course, clothing. Social change usually results in the disappearance or diminution of local styles and the adoption of the more widespread "modern" dress.

It is the authors' hypothesis that children who, when asked to draw a man, draw chiefly men in modern dress show thereby a preference for modern dress and for the complex of social changes of which it is a part. On the other hand, children who for the most part draw men having the traditional appearance of men of their group are either (a) unacquainted with modern dress or (b) hold negative attitudes toward it.

Unpublished data obtained by Dennis suggest that changes in human-figure drawings occur early in the process of modernization. This is what one would expect if, as has been suggested, psychological acceptance must precede the modernization that is sponsored by the changing group itself.

For example, unpublished data from Lebanon show that few men in traditional dress are drawn by children in villages in which traditional dress is still worn by the older men. The same results have been obtained from Indian groups in Mexico. These results suggest that drawings obtained from countries which have been less exposed to modern influences than have been Lebanon and Mexico would be of interest.

The present paper, as the title indicates, is concerned with drawings obtained in Sudan, a country many parts of which meet the requirement proposed above. The data were gathered by one of the authors (Badri) who is Sudanese and who is familiar with the schools and the communities in which drawings were obtained. The assistance of several Sudanese students at the American University of Beirut and at the Ahfad College of Omdurman, Sudan, is gratefully acknowledged.

#### B. PROCEDURE IN OBTAINING DRAWINGS

All drawings were obtained by Dr. Badri or by a trained assistant. They were made in the children's usual classrooms. They were drawn with pencil on plain white 9" × 12" paper. The children were asked to draw a man, any kind of man they wished, but to draw a whole man—not just a head or torso. There was no time limit.

The test was given in Grade 4, in which the pupils were primarily 10 and 11 years of age. This grade was chosen because children draw a man less well before ages 10 and 11 and because a considerable degree of elimination from school occurs beyond these ages.

### C. THE GROUPS TESTED

Data are presented only for boys. Drawings were obtained from four boys' schools in Omdurman and Khartoum, the two major cities in Sudan. In addition, a boys' school in each of two villages was tested.

In terms of modernization of the families of the students, Dr. Badri and several Sudanese university students ranked these schools as follows (prior to classification of the drawings): Rated as most modern were the families of students in an American school for boys which has been operated in Omdurman for some time. While all of its teachers and students are Sudanese and instruction is in Arabic, by tradition it is "modern" and parents who wish their sons to have this sort of outlook send them to this school. This school was rated "A" in respect to modernization.

The other Omdurman and Khartoum schools which were tested were rated "B." They have been combined for the purposes of this paper. It was unanimously agreed by the raters that both villages rank below the urban groups and that village "D" is below village "C." Village C has a large center for religious training and for native psychotherapy. This village has electricity in the religious center and is situated on a main traffic highway. Because village C is nearer to Khartoum, its people have more opportunity to visit the capital. Village D is farther away from the capital and is much smaller. It has no electricity, its houses are poorer, its people are poorer. It has poor communication with the capital.

### D. THE CLASSIFICATION OF THE DRAWINGS

In the case of each drawing it was determined by inspection whether it represented (a) a man of traditional appearance, (b) a man in modern dress, or (c) a man of uncertain or ambiguous appearance. The decisions depended primarily upon clothing, but other items sometimes played a part.

The most distinguishing aspect of the appearance of a traditional Sudanese man, as depicted in the drawings, is his chief article of apparel: the galabia. This is a long loose garment extending from the shoulders to the ankles. To a person accustomed to modern dress, it is likely to suggest a nightgown. It is made of cotton and ordinarily is white. Because of its simplicity, it is easily drawn and is not difficult to distinguish from drawings of modern dress.

Under the galabia are worn thin cotton trousers resembling pajamas. Children in their drawings frequently draw these trousers although they are, in fact, not visible. Shoes without laces may be worn with the galabia, but the poorer Sudanese wear a native type of slipper.

The traditional Sudanese dress includes several forms of headgear, which



need not be described here, but traditional Sudanese headgear is easily distinguished from modern hats and caps. Formerly many Sudanese had their cheeks scarified, but scarification of the face is seldom represented in the drawings. Apparently the abandonment of facial scarification is one of the first steps toward modernization.

In Sudanese drawings, modern dress can be distinguished from traditional dress by the presence of one or more of the following items: a hat or cap, collar, necktie, shirt, coat, belt, and trousers not accompanied by a galabia.

In "good" drawings the distinction between traditional and modern appearance is easily made. However, many of the drawings are not good. In poor drawings, if the drawing contained a mixture of modern and traditional, such as a galabia combined with a wristwatch, it was classified according to the aspect of appearance that seemed to predominate: in this case, the galabia. Some drawings, however, were not classifiable, either because they represented unclothed figures or were so poorly executed that they could not be interpreted or because they appeared to represent a man neither in current traditional dress nor in current modern dress. As was noted earlier, we have therefore three classes of drawings: (a) traditional, (b) modern, and (c) neither traditional nor modern.

After the non-Sudanese author (Dennis) had been familiarized with the drawings and with the costumes which they represented, the two authors independently scored 100 papers chosen from the four ranks. The agreement with respect to placement in the three categories was 98 per cent. The data included in Table 1 were classified by Badri and by Sudanese assistants whose independent scoring also agreed highly with that of Dr. Badri.

### E. RESULTS

The results to be considered are shown in Table 1. In this table the successive columns show for each group (a) the number of drawings, (b) the number which were classified as traditional, (c) the number classified as modern, and (d) the per cent of the combined traditional and modern drawings which were modern.

The results require little comment. The per cents of modern drawings parallel the group rankings for modernization. The per cents of modern drawings vary from 82.9 in the most modernized group to 4.9 in the least modernized. If each group is compared with all other groups, in all such comparisons the group designated as more advanced in modernization has the higher per cent of modern drawings. In all such comparisons, the differences are significant at the .001 level of confidence.

TABLE 1  
NUMBER AND PER CENT OF TRADITIONAL AND MODERN DRAWINGS  
BY SUDANESE CHILDREN

Group	Number of cases	Drawings		Per cent modern
		Traditional	Modern	
A	41	6	29	82.9
B	178	86	64	42.6
C	28	19	5	20.8
D	46	39	2	4.9

The fact that modern dress is seldom drawn in the two villages is not due to lack of acquaintance with it. The teachers of these boys are men in modern dress. The drawing test was administered by a psychologist in modern dress. Other men in modern dress, such as policemen, agricultural officers, and health officers, either live in the village in question or frequently visit it. However, the prominent local citizens usually wear traditional dress. The data, therefore, are consonant with the interpretation that most of the village boys and some of the urban boys have not yet developed a preference for modern clothing, although it is familiar to them. It is predicted by the authors that before these villages in Sudan adopt the technological changes associated with modernization, the content of drawings will begin to show a preference for modern dress.

#### F. SUMMARY

It is proposed that the relative frequencies of traditional appearance and modern appearance in children's drawings of a man made by children who are familiar with both kinds of dress reflect the extent to which the child favors one form of dress over the other. It is further proposed that in Sudan, where some groups are much more "modernized" than others, the degree of adoption of the modern style of life would be reflected by the relative frequency of modern dress in the drawings of various groups. Drawings were obtained from Sudanese groups appropriate for the study. The findings are consonant with the interpretation proposed.

#### REFERENCE

1. DENNIS, W. Values Expressed in Children's Drawings. In W. Dennis (Ed.), *Readings in Child Psychology* (2nd ed.) Engelwood-Cliffs, N.J.: Prentice-Hall, 1963. Pp. 265-271.

Department of Psychology  
American University of Beirut  
Beirut, Lebanon

Department of Psychology  
Brooklyn College  
Brooklyn 10, New York

## THE MOOD EFFECTS OF MUSIC: A COMPARISON OF DATA FROM FOUR INVESTIGATORS\*

*Department of Psychology, Wisconsin State University—Eau Claire*

---

MELVIN G. RIGG

---

### A. INTRODUCTION

What is the meaning of a piece of music? To the professional musician, significance is apt to be placed upon intrinsic meanings embodied in harmonic, contrapuntal, or formal relationships. On the other hand, the musical layman often seeks for extrinsic meanings beyond the music. The extrinsic meanings may be divided into two subtypes: (a) mood effects and (b) definite ideas and actions that the music is supposed to portray. Works characterized by the latter have been given the designation "program music."

The assumptions of program music have failed to stand up under experimental attack. Of course, if one is told beforehand what a particular selection is supposed to represent, it is easy to imagine that the description is valid; but when without such knowledge the attempt is made to judge unfamiliar music, the situation is very different. Thus in 1929 Wells (24) demonstrated that listeners were usually unable to match compositions with descriptive titles given in another order, and this result was confirmed by Rigg (14) in 1937.

Music, however, does possess mood effects; and these effects have been studied by measuring physiological reactions and by verbal report. Music can produce changes in breathing, pulse rate, blood pressure, and galvanic response; but the attempt to elaborate a theory based upon such data has proved disappointing. The verbal report, in spite of its subjectivity, has been more satisfactory than the physiological measurements; and many experimenters, among them Schoen and Gatewood (20, 21), Campbell (1), Hampton (7), and Dreher (3)—in addition to Kate Hevner Mueller, whose work will be described later—have shown that there is considerable consistency in the reports given for certain compositions.

Best results have been secured not when the auditors are asked, "How does this music make you feel?" but simply, "What mood does this music express?" since the former question involves a reaction with the existing

---

\* Received in the Editorial Office on July 9, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

moods of the listeners. If one is depressed, music may not make him happy; but he may still recognize some music as "happy" music.

Mood effects are not to be found in any great degree in individual notes, intervals, or chords, for one chord constitutes music no more than one word constitutes a poem. Complete portions of music are also better than unaccompanied melodies. On the other hand, the attempt to characterize an entire composition often has resulted in confusion since the mood may change from section to section. For an interesting example, see Hevner Mueller's report (11) on her results with Debussy's "Reflections on the Water." Conflicting opinions were resolved when three portions of that composition were judged separately. The best results are secured from very short bits of complete music, ranging from a four-measure phrase to perhaps a two-minute passage.

### B. THE ADJECTIVE CLOCK

The use of verbal reports involves a problem in semantics. If listeners are allowed to record anything that occurs to them, the results are apt to be so varied as to make it difficult to draw conclusions. To obviate this difficulty, checklists frequently are provided. Checklists, of course, have a suggestive effect and may modify the results. The most interesting checklist is the Hevner Mueller devised "adjective clock" consisting of 67 words grouped into eight categories (11). In each group the words are supposed to represent similar moods, and adjacent clusters are supposed to deviate slightly by cumulative steps until groups at opposite points are in contrast. The deviation continues until Group VIII, which is similar to Group I. The general nature of this clock is shown in the abbreviated diagram in Figure 1.

Several investigators have suggested shifts in this arrangement, but these shifts usually have involved changes between adjacent groups only. On the basis of numerous correlations between the terms, Farnsworth (4) has recommended nine clusters. Farnsworth's evidence supports the homogeneity within each cluster to a greater extent than it supports the clocklike cumulative relationship of the groups. His new suggested arrangement improves the situation, but does not entirely solve the problem because there still remains in one instance the low correlation of .14 between adjacent clusters.

In a doctoral dissertation at Duke University, Watson (23) reports 15 moods that can be discriminated. In many cases, Watson has split the Hevner Mueller groups into two subgroups (in one case into three subgroups), while certain other groups have remained the same. In short, discrepancies in the use of terms are not sufficiently serious to prevent a comparison of the results of various research workers.

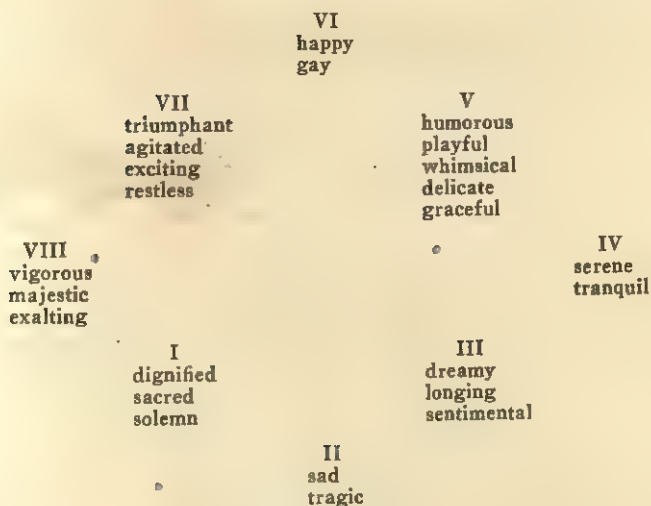


FIGURE 1  
HEVNER MUELLER'S ADJECTIVE CLOCK

Several experimenters have gone beyond the point of asking merely what mood a piece of music represents and have endeavored to ascertain just what features of the music account for the mood effect. To those who may view this field as one capable of yielding only nebulous results, it may come as a surprise to find a general agreement. Tables 1 to 8 present the data of four investigators: Kate Hevner Mueller (9, 10, 11, 12), Watson (23), Gundlach (5, 6), and Rigg (14, 15, 16, 17, 18, 19). Each table presents characteristics of music found by Hevner Mueller to be associated with one of the clusters in her adjective clock, together with the data of the other workers for the same or similar moods expressed in terms that they have used.

### C. DISCUSSION OF THE TABLES

Table 1 indicates a general agreement that music that is characterized as dignified, serious, solemn, or sacred is slow, low-pitched, and avoids irregularity of rhythm and dissonant harmonies.

Table 2 shows that sad music is low-pitched, slow, apt to be in the minor mode, and apt to contain dissonance. Certain differences appear between the sad and the tragic, as well as between melancholy music and the music of lamentation.

Sentimental or pleading music (see Table 3) is apt to be slow and high-pitched. This category appears less definite than those given in Tables 1



TABLE 1  
GROUP I FROM HEVNER MUELLER'S ADJECTIVE CLOCK

Musical feature	Hevner Mueller dignified sacred solemn	Watson	Watson	Gundlach	Rigg
rhythm	firm (hymn style)	dignified	serious	dignified	solemn
phrasing		regular	regular		trochaic
tempo	slow	moderately slow	slow	slow	legato slow
register	low	moderately low	low	moderately low	low
mode	major				major
melody	ascending			narrow range	
harmony	simple	simple	simple		simple
dynamics		few changes	few changes moderately soft		

TABLE 2  
GROUP II FROM HEVNER MUELLER'S ADJECTIVE CLOCK

Musical feature	Hevner Mueller sad tragic	Watson	Watson	Gundlach melancholy mournful	Rigg	Rigg
rhythm	firm (hymn style)	moderately regular	moderately irregular			
phrasing					legato	legato
tempo	slow	slow	moderately slow	moderately slow	moderately slow	slow
register	low	low	moderately low	low	moderately low	low
mode	minor					minor
melody				narrow range		
harmony	complex	complex	complex			complex
dynamics		few changes moderately soft	moderately loud	soft		

TABLE 3  
GROUP III FROM HEVNER MUELLER'S ADJECTIVE CLOCK

	Hevner Mueller	Watson	Gundlach	Rigg
Musical feature	dreamy longing sentimental	pleading	sentimental	sorrowful longing
rhythm	flowing			
phrasing	"		"	legato
tempo	slow		slow	slow
register	high	high	high	
mode	major			
melody	ascending		narrow range	
harmony	simple			
dynamics		quick changes  moderately loud		

TABLE 4  
GROUP IV FROM HEVNER MUELLER'S ADJECTIVE CLOCK

	Hevner Mueller	Watson	Gundlach
Musical feature	serene tranquil	peaceful dreamy calm	tranquil
rhythm	flowing	regular	"
tempo	slow	moderately slow	slow
register	high	"	low
mode	major		"
melody	ascending		narrow range
harmony	simple	moderately simple	
dynamics		no changes  soft	moderately soft

TABLE 5  
GROUP V FROM HEVNER MUELLER'S ADJECTIVE CLOCK

Musical feature	Hevner Mueller humorous playful whimsical delicate graceful	Watson mischievous playful	Watson amusing funny	Gundlach whimsical	Gundlach flippant	Gundlach delicate
rhythm	flowing		irregular		smooth	
tempo	fast	fast	moderately fast	fast	fast	moderately slow
register	high	moderately high	moderately high	moderately high		
mode	major					
melody	descending			wide range		narrow range
harmony	simple	moderately simple				
dynamics		quick changes	quick changes			soft

TABLE 6  
GROUP VI FROM HEVNER MUELLER'S ADJECTIVE CLOCK

Musical feature	Hevner Mueller happy gay	Watson happy glad	Watson very happy exuberant	Gundlach glad	Rigg gaiety	Rigg joy
rhythm	flowing	regular		smooth	trochaic	iambic
phrasing					staccato notes	
tempo	fast		fast	fast	fast	fast
register	high	moderately high	moderately high	moderately high		high
mode	major					major
melody				wide range		
harmony	simple	simple	simple			simple
dynamics		moderately few changes	some changes		moderately soft	loud
			moderately loud			

TABLE 7  
GROUP VII FROM HEVNER MUELLER'S ADJECTIVE CLOCK

	Hevner Mueller	Watson	Watson	Gundlach	Gundlach	Rigg	Rigg
Musical feature	triumphant agitated restless	exciting	very exciting	uneasy	triumphant	agitation	triumph
rhythm	firm (hymn style)			rough			
phrasing							
tempo	fast	fast	fast	fast	moderately fast	staccato notes fast	fast
register	low	high	high			low	high
mode						minor	
melody	descending			wide range	narrow range		
harmony	complex		complex	moderately loud	loud	complex	
dynamics		loud	some changes loud				

TABLE 8  
GROUP VIII FROM HEVNER MUELLER'S ADJECTIVE CLOCK

Musical feature	Hevner Mueller vigorous majestic exalting	Watson  majestic	Gundlach  exalted
rhythm	firm (hymn style)	regular	
tempo	fast		slow
register	low		
melody	descending		
harmony	complex	simple	
dynamics		moderately few changes moderately loud	moderately loud

and 2. In Farnsworth's rearrangement of the adjective clock, "dreamy" and "sentimental" are placed in one group, while "longing" and "pleading" are placed in another.

Calm or dreamy music is apt to be slow and soft and is apt to contain little dissonance (see Table 4). Farnsworth places "dreamy" in the same cluster as "serene" and "tranquil," although Kate Hevner Mueller separates it from the latter two.

Graceful or playful music, according to Table 5, is high-pitched and fast. There is some question whether "delicate" belongs in this group. Farnsworth splits Kate Hevner Mueller's Group V into three separate clusters, putting "playful" in Cluster A, "whimsical" in Cluster B, and "delicate" and "graceful" in Cluster C. "Humorous" falls between three clusters, A, B, and I, but is closest to A.

Table 6 shows happy music to be high-pitched, fast, to contain little dissonance, and to be in the major mode. Very happy or joyous selections are loud.

Triumphant or exciting music (see Table 7) is fast, loud, and is apt to contain dissonance. To the writer, it seems a mistake to put "agitated" and "triumphant" together because the former is unpleasantly toned while the latter is "happy." Farnsworth separates the two terms and in fact divides Hevner Mueller's Group VII between two clusters. In the writer's own experiments, a most startling result was the production of the effect of agitation when staccato notes were interjected into a sorrow pattern. None of the other investigators has considered this phenomenon.



According to Table 8, exalted or majestic music seems to involve a firm or regular rhythm and to be moderately loud. Perhaps there is less agreement here than in the other tables. Farnsworth divides Hevner Mueller's Group VIII between the same two clusters into which he divided Hevner Mueller's Group VII. Evidently semantic difficulties make trouble at this point.

As a general summary of Tables 1 to 8, it can be said that the points of agreement far outnumber the divergences. Some of the latter are due to problems of grouping terms.

#### D. RESULTS FROM THE STANDPOINT OF THE MUSICAL FEATURES

This treatment is the reverse of that of the tables in that the various musical features will be discussed in order.

Of the various characteristics, tempo (according to Hevner Mueller, Gundlach, and Rigg) is the most important feature in determining the mood effects of music. Other things being equal, a fast speed has an effect in the direction of joy, whereas a slow speed tends toward sorrow. This finding is also supported by Bennett Shimp (22). It must be emphasized, however, that the mood effect is the result of a total pattern. Not every fast piece is joyful, but as a fast piece of music is speeded up it becomes more joyful or at least less sorrowful. The clearest proof for this comes from Rigg (18) in an experiment in which each of five phrases having different mood effects was played at six different metronome speeds. When played fast, a phrase having sorrowful effects may suggest agitation (16). If other features permit, very fast music is "very exciting" (23) or "uneasy" (6). On the other hand, slow music may be sad, calm, or dignified, depending on other features (6, 12, 15, 16, 18, 19).

Loud music is apt to be very happy, exciting, or triumphant (6, 23). Watson (23) says that when both loudness and tempo are increased, happy music becomes "very happy," whereas sorrowful music changes to "tragic." Soft music is apt to be sad, serious, peaceful, or delicate (6, 23). According to Rigg (16), when a phrase suggesting joy was reduced in loudness, it was interpreted as "gaiety," defined in his experiment as a lesser degree of joy. According to Watson (23), quick changes in loudness suggest the very happy or the very exciting, whereas few or no such changes are found in peaceful, dignified, and sad music.

High pitch suggests the happy mood (6, 12, 17, 23). Low pitches are apt to be dignified, serious, or sad. If other features permit, a low register may suggest the tragic mood.

Melodic range has been studied only by Gundlach (6). He found a wide

range in happy, whimsical, or uneasy music; a narrow range in dignified, sad, sentimental, tranquil, delicate, or triumphant selections.

Hevner Mueller (11) and Rigg (16) agree that the rise and fall of melody are not important ordinarily.

Heinlein (8) has pointed out that a major passage may be sad (Dvorak's "Largo") and that a minor composition may be cheerful ("Anitra's Dance" by Grieg). However, other things being equal, the major mode has an influence in the direction of the happy, while the minor mode tends toward the sad. These mode effects have been studied by Hevner Mueller (9) and Rigg (16).

Simple harmonies indicate the happy, graceful, and serene. Rigg (16) found that a phrase suggesting joy was less joyful when more complex chords were introduced. Hevner Mueller (11) and Watson (23) agree that dissonance or complex harmony produces sad, tragic, or very exciting effects.

Irregular or rough rhythms suggest the tragic or amusing (23) or the uneasy (6). For Watson (23) regular rhythms characterize dignified, serious, sad, peaceful, happy, or majestic music (differences among these moods depending on features other than rhythm). Gundlach (6) assigns regular rhythms to happy or flippant music.

According to Hevner Mueller (11), passages in hymn style (that is, passages consisting of chords) suggest the sad, dignified, vigorous, or triumphant; flowing melodies suggest the happy, graceful, dreamy, or peaceful mood.

Gundlach (5) noticed that in songs of American Indians a rhythmic staccato indicated excitement. Rigg (16) found that staccato notes introduced into a phrase otherwise sorrowful produced agitation; introduced into a joyful phrase, they produced the effect of gaiety.

### E. GENERAL DISCUSSION

What effect does musical training have upon mood discrimination? Watson (23) found that, while lower levels of musical training have no effect, highly trained musicians are better than untrained persons. However, some musicians make low scores on the Watson test for mood discrimination; and some high scores are made by persons without musical training. Hevner Mueller states that only to a limited degree are trained musicians better than untrained persons at discriminating between major- and minor-mood effects. Watson (23) and Hevner Mueller (9) agree that mood discrimination has little relationship to intelligence or to scores on the Seashore Measures of Musical Talent. Watson says, however, that those who have a high appreciation of mood effects are those who enjoy music.

Are the mood effects discussed in this paper peculiar to a Western European heritage or do they have a wider application? Gundlach (5), in his analysis of songs of American Indians, has made an interesting attack on this question and assumes that the effects transcend cultural differences; but Morey (13) reports that German music expressing love and fear did not express emotion to Liberians. This sort of investigation should be continued.

In conclusion, it can be said that the mood effects of music are well established and that there is a high level of agreement with regard to the features that are responsible. The evidence does not, however, justify the claims made for program music when such music is said to represent definite ideas and actions.

#### REFERENCES

1. CAMPBELL, I. G. Basal emotional patterns expressible in music. *Amer. J. Psychol.*, 1942, **55**, 1-17.
2. DOWNEY, J. E. A musical experiment. *Amer. J. Psychol.*, 1897, **9**, 63-69.
3. DREHER, R. E. The relationship between verbal reports and galvanic skin responses to music. Doctor's thesis, Indiana University, Bloomington, Indiana, 1947.
4. FARNSWORTH, P. R. A study of the Hevner adjective list. *J. Aesth. & Art Criticism*, 1954, **13**, 97-103.
5. GUNDLACH, R. H. A quantitative analysis of Indian music. *Amer. J. Psychol.*, 1932, **44**, 133-145.
6. ———. Factors determining the characterization of musical phrases. *Amer. J. Psychol.*, 1935, **47**, 624-643.
7. HAMPTON, P. J. The emotional element in music. *J. Gen. Psychol.*, 1945, **33**, 237-250.
8. HEINLEIN, C. P. The affective characters of the major and minor modes in music. *J. Comp. Psychol.*, 1928, **8**, 101-142.
9. HEVNER, K. (Mrs. J. H. Mueller). The affective character of the major and minor modes in music. *Amer. J. Psychol.*, 1935, **47**, 103-118.
10. ———. Expression in music: A discussion of experimental studies and theories. *Psychol. Rev.*, 1935, **42**, 186-204.
11. ———. Experimental studies of the elements of expression in music. *Amer. J. Psychol.*, 1936, **48**, 246-268.
12. ———. The affective value of pitch and tempo in music. *Amer. J. Psychol.*, 1937, **49**, 621-630.
13. MOREY, R. Upset in emotions. *J. Soc. Psychol.*, 1940, **12**, 333-356.
14. RIGG, M. G. An experiment to determine how accurately college students can interpret the intended meanings of musical compositions. *J. Exper. Psychol.*, 1937, **21**, 223-229.
15. ———. Musical expression: An investigation of the theories of Erich Sorantin. *J. Exper. Psychol.*, 1937, **21**, 442-445.
16. ———. What features of a musical phrase have emotional suggestiveness? Report No. 1, Social Science Research Council of the Oklahoma A. & M. College, 1939.
17. ———. The effect of register and tonality upon musical mood. *J. Musicol.*, 1940, **2**, 49-61.

18. ———. Speed as a determiner of musical mood. *J. Exper. Psychol.*, 1940, **27**, 566-571.
19. ———. The expression of meanings and emotions in music. In F. P. Clarke & M. C. Nahm (Eds.), *Philosophical Essays in Honor of Edgar A. Singer, Jr.* Philadelphia: Univ. Pennsylvania Press, 1942. Pp. 279-294.
20. SCHOEN, M., & GATEWOOD, E. L. The mood effects of music. In M. Schoen (Ed.), *The Effects of Music* (Chap. 7). New York: Harcourt, Brace, 1927.
21. ———. Problems related to the mood effects of music. In M. Schoen (Ed.), *The Effects of Music* (Chap. 8). New York: Harcourt, Brace, 1927.
22. SHIMP, B. Reliability of associations of known and unknown melodic phrases with words denoting states of feeling. *J. Musicol.*, 1940, **1**, 22-35.
23. WATSON, K. B. The nature and measurement of musical meanings. *Psychol. Monog.*, 1942, **54**, Whole No. 244.
24. WELLS, F. L. Musical symbolism. *J. Abn. & Soc. Psychol.*, 1929, **24**, 74-76.

*Department of Psychology*  
*Wisconsin State University*  
*Eau Claire, Wisconsin 54701*

## AN INDIVIDUAL ACTION-RESEARCH METHOD FOR INTENSIVE EXPLORATION OF ADAPTIVE DIFFICULTIES IN THE PREPUBERTY STAGE\*<sup>1</sup>

*The Menninger Foundation*

---

C. J. MORGAN

---

### A. INTRODUCTION

This is a report on a way of using activities for increased understanding of a boy's problems, and for discovery of his potentialities, at a transitional stage when both internal conflicts and difficulties with the environment are seriously blocking his progress.

This project represents one focus of a larger study of prepuberty adjustment and the personality development of 60 originally normal children. The larger study has as its concern the understanding of the dynamics of the developmental process at puberty. Accordingly, the emphasis is upon individual differences in different family settings and upon the different kinds of internalizations and representations of sequelae of previous experience as these contribute to problems in dealing with the environment or to solutions of these problems. A related aim has been to explore the range of resources that a given child can use at the prepuberty stage and the extent to which the provision of resources relevant to the child's interests can assist in the better understanding of the child's motivations, conflicts, and needs. In addition, we sought to assess how the provision of relevant resources could assist a child in dealing with his problems with the environment and within himself.

Preadolescent boys showing moderate to severe difficulties with peers and in school activities were chosen for intensive probing study, through specially planned experiences. The case of one boy, Peter, is selected for review here. Peter does not represent the most dramatic example of the success of this approach in evoking the potentialities of the boy, but rather the one into which has gone to date the greatest effort to evaluate the method in terms

---

\* Received in the Editorial Office on July 9, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This project was carried out as part of a group of studies in Maintenance of Mental Health at Prepuberty, USPHS Grant 4093, L. B. Murphy, Chief Investigator. Subjects were selected as infants (to participate in the study of infancy by Escalona and Leitch) and were regarded as normal in the sense that they seemed to be developing, at least at an average rate, and were in good health and free from defect.



of its usefulness as a research technique. This preliminary report illustrates only a limited part of the findings from this study.

## B. BACKGROUND

At the outset, Peter, a tall but chubby 12-year-old with bushy dark-brown hair, alert eyes, and a complexion as fair as that of a girl, was selected because of the great discrepancy between his excellent preschool and his poor current coping techniques and his inadequate level of functioning in all aspects of his preadolescent life. At 4 years of age he was described as follows (in a structured test session):

Peter appeared to be a happy, comfortable child whose coping style was generally direct and appropriate for problem solving, as well as self-gratifying. He has a great deal of self-respect and realistically based self-confidence which allowed him to view himself and his world with a positive expectation of doing well and being liked. Although he recognized that there were things he could not do, he was rarely discouraged and could resiliently slough off temporary disappointment or difficulty. One might say that he enjoyed and used his intellectual resources for immediate and realistic problem solving without pressure to achieve any distant or long-standing goals (Moriarty).

During prepuberty Peter revealed notable changes, as was pointed out by the examiner again in a structured testing situation.

At the outset, Peter seemed subdued. He smiled readily, but he was easily embarrassed, often flushing for no apparent reason. It seemed as though he complied with test demands out of courtesy rather than interest. He was quick to understand and to respond, but lack of elaboration suggested superficial effort and a lazy stolidity which rarely permitted concentrated reflection. Mediocrity appeared in Peter's incomplete sentences, use of trite colloquial phrases, easy satisfaction with approximations and inability or unwillingness to press for clarity which was not obvious. Whether the present achievement drive is a permanent adjustment to relatively low family standards or a temporary aspect of peer adjustment is not known on the basis of these tests (Moriarty).

This gross contrast between Peter's preschool and present levels of flexibility, constructiveness, vigor, and achievement drive did not seem to be sufficiently accounted for by existing test records and reports. Also, we did not know how his present difficulties were affecting his current performance in school and his behavior at home. Therefore, Peter's teachers were interviewed and his behavior was observed in school and at home.

An interview with Peter's sixth-grade teacher and the school principal during February of 1961 revealed that he was engaging in frequent severe

fight, and the principal considered him a "borderline delinquent." An interview with his teacher revealed that many of his difficulties resulted from a strong need to be first or best in everything without effort; if he could not be the best or first he either would not participate or he would cheat to obtain his goal. He refused to participate in organized sports, preferring instead free-play games, such as kickball, which he played quite well. In the classroom little evidence was revealed of his superior intellectual capacity, as his grades averaged Ds and Cs.

Observations of Peter at home revealed additional evidence of his aggression as he told of setting his cousin's pants on fire, of pulling Halloween costumes off younger children and taking their candy, and of beating upon his younger brothers and sisters when he was angry or disappointed.

Helping to provide Peter with undesirable outlets for his aggression was Ned, another bright, energetic boy, who lived next door to Peter and who (with his variety of ideas) was admired by Peter. Ned was often the source of escapades that led Peter into difficulty. For example, it was Ned who taught Peter to smoke grapevines and to make small pistols which would fire a single buckshot.

The observations at Peter's home made it apparent that Peter felt that he was entitled to whatever he wanted and that anything that came his way was a gift which it was not necessary to repay. In fact, he seemed to regard his friends as fools for buying him things, as though they just could not keep from giving to him.

The ease with which Peter and Ned (during the observations at Peter's home) related their feelings, attitudes, and escapades to a young male graduate student led to a later consideration of the possibility that, perhaps with additional sessions, more could be learned about Peter.

While all intelligence tests showed his superior capacity (*IQ* average = 125), Peter's grades, averaging D+ to C—, indicated that he was not able to use (or did not use) much of his ability. His desire to be best, and the aggressive behavior in which he engaged, interfered with his school work as well as with his relationships on the playground.

Peter rejected Boy Scouts because, as he put it, "We never did anything." Working on merit badges for which he could see no use seemed like nonsense to him. Peter held a membership in the local YMCA and made frequent use of it for swimming, but he exerted only minimal effort in improving this skill.

At the time that Peter's difficulties led to this intensive study, the reason

for increasing resistance to school responsibilities were far from clear; nor was it clear why he had become so destructively aggressive—hence it was decided to supplement test evaluations by selected activities that would permit more detailed assessment of the meaning of his behavior and that could also explore possible avenues for change. Detailed process records of each session were dictated immediately; excerpts from these will be used to document the major points of our learning about Peter.

### C. ACTION-RESEARCH PROGRAM

This study was viewed initially as an action-research approach to the more complete understanding of a boy in difficulty. The decision to investigate Peter's present difficulties (through one of prepubescent boys' major avenues of interaction with their environment, that of action in activities) led to considering this technique as an action-research approach. With this approach, we hoped to explore (through activities and opportunities that we have called "Growth Experiences") the attitudes, feelings, conflicts, and obstacles that underlay the disturbed functioning and, simultaneously, the capacity of a boy to respond to opportunities which would build on our discovery of his needs.

A few months after the home observations, Peter and his younger brother Rex were approached by the young male graduate student who had done the previous observing. The boys were asked if they would like to participate in another part of the child-study project with which they had been involved since infancy, a part that would focus upon what kind of activities boys their age enjoy. Both Peter and Rex seemed interested, but they wanted to know what kinds of things they would be doing. The young male graduate student, who for the sake of brevity will be referred to as the GEP (Growth Experience Provider), suggested that he plan some activities for the first session and then in following sessions the decision as to what activities they would engage in would be up to the boys. Peter and Rex, realizing that they did not have to take part in this study and that if they lost interest in it they were free to quit, willingly accepted the GEP's offer.

#### *1. Initial Growth Experiences: The Use of Slings To Test Peter's Capacity To Shift to Socially Acceptable Outlets for Aggression*

The growth experiences were initiated with the construction of slings, as used by David in the Bible, partly because of the strength of Peter's aggressive drives and partly to help him break his dependence on Ben and Ben's aggressive and destructive ideas. The GEP's notes read as follows:

The GEP asked Peter and Rex if they had ever made a sling. They did not know what a sling was; so the GEP stated that he was referring to a sling of the type used by David to kill Goliath, as related in the Bible. Chester remarked that he knew the story but had never made such a sling.

The GEP then constructed a sling from the tongue of an old shoe and two old leather boot laces. He demonstrated how the boot straps provide added leverage to the arm and increase the distance and speed with which an object can be thrown. The boys were amazed.

Peter was then given the sling and shown how to wrap one of the straps around his wrist while he held the other one in his hand. A rock was placed in the cup formed by the shoe tongue, and then he was told to swing it around several times and then to let loose of the strap he held in his hand.

Peter *practiced with the sling* and with different sizes of rock and improved rapidly.

He *discovered* that a small very round rock worked best and that he could increase the distance that he could sling a rock by untying the strap around his wrist, making a slipknot in the end of it, and then slipping it over one finger. This technique which Peter devised gave him six to eight additional inches of strap and was a nice adaptation of the principle of leverage which the GEP had stated to him.

Thus, we learned that Peter, outside of the competitive school situation and in an atmosphere of being helped by a friendly older male, had a capacity to practice. He also showed, from the first, real ingenuity and initiative in solving problems which he met in pursuing an interest of his own.

Both boys were enthusiastic about the slings and stated that they were going to make one just as soon as they reached home.

A week later the GEP arrived at their home, and Peter hurried out to his car and asked if the GEP had brought his sling. Rex joined Peter and the boys said that they wanted to go out to the lake and try out their slings. Both boys had slings which, the GEP learned later, Peter had made. Peter had also made a sling for almost every boy in the neighborhood.

Rex mentioned that Peter was getting quite good at aiming with his sling. But he also mentioned that some of the boys in the neighborhood were using the slings to throw rocks at each other, and Peter was not an exception.

At the lake Peter showed considerable skill with his sling, which undoubtedly resulted from practicing (as Rex had mentioned). The day was quite cold and yet both boys, particularly Peter, wanted to continue practicing with their slings. The fact that Peter would practice on something at which he initially had no skill was a contradiction to a statement made by Peter's teacher concerning Peter's inability to practice anything in which he lacked some skill.



## 2. *Swimming Used To Explore Problems in Competition and Fear of Aggression from an Adult*

At the beginning, swimming was also explored as an outlet for aggression and as a source of just playful fun. But, in the pool, Peter learned for perhaps the first time that he had to exert himself to win a race: i.e., to be first. He was taught new swimming strokes and was encouraged to practice them by working them into games. It was also in the pool that Peter was encouraged to be competitive with the GEP. At first he had to be urged and taught how to duck the male adult. Sometimes he even had to be helped to do this but, from this dependent beginning, he developed to the point where he and his brother planned strategies to duck the GEP and were able to do so spontaneously. From this sequence and from the simple feat to which it led, we learned how great had been Peter's fear of competition with a male adult. Here he first began to overcome this fear of outdoing an adult and of retaliation from an adult. The swimming also served as the first place where Peter learned to compete when he had no chance of winning, as in a race with the GEP, but this development came slowly.

During the first session with Peter and Rex, the GEP took the boys swimming at the local YMCA. Both boys could swim, but Peter was much more skillful than Rex. The GEP showed Peter a couple of new strokes, which Peter practiced on for only a few minutes. He was not interested in making this a period in which he had to learn to improve some skill.

In an attempt to teach the boys how to swim under water, the GEP and the boys threw a penny on the bottom of the pool. Peter made two or three attempts to find the penny, but was unsuccessful. He decided that it was not worth the effort; so he sat and watched Rex and the GEP dive for it.

About three weeks later, after Peter and Rex had played around in the pool for a few minutes, Peter asked the GEP to show him the frog kick again. Peter had little difficulty in learning a new stroke, but was not very interested in practicing. He usually would practice a very short time and stop. But soon he practiced longer and, with encouragement from the GEP, made more efforts to improve.

Peter frequently ducked Rex and, since he is so much larger than Rex, Rex found it difficult to duck Peter. The GEP then showed Rex an easy way to duck anyone, regardless of size. Peter sat attentive on the edge of the pool and was very attentive to the instructions given to Rex on ducking. When the GEP was not looking, Peter swam up behind him and ducked the GEP—using the technique that had been demonstrated to Rex.

Later a lifesaving class entered the pool, and the instructor began explaining how to handle a drowning person's attempts to grasp the



lifesaver. Peter listened to the instructions and watched the demonstration, but did not understand the need to be instructed in how to handle a drowning person nor why such a person would grab the lifesaver. After a brief discussion Peter understood that, if someone grabbed him, he would have trouble swimming. The GEP then had Peter swim with him into deep water and instructed Peter to try to grab him and hang on. As Peter tried with his right hand to grab the GEP, the GEP (with his right hand) grabbed Peter's arm and pulled Peter toward him, thus turning Peter around. Then he reached across Peter's chest and grabbed him under the arm. Peter was surprised, but liked this and asked to see it again. Peter was shown a couple more times and then allowed to try this technique on the GEP.

Afterwards Peter asked the GEP where he learned to swim so well. Peter had forgotten that this had been explained to him before. Also, Peter seemed to feel that the GEP had become an experienced swimmer overnight. The GEP then told Peter again about the years of lessons and practice, the steps in the development of this skill, and what the GEP had accomplished at what ages.

The next week Peter suggested that he and Rex have some races to see who could swim the farthest under water. Rex won the first race, which annoyed Peter; so he said, "I'll race you again." Peter said, "This time, let's see who can go the farther without staying under water." This time Peter barely won the race. On the next race Peter went farther, but cheated and therefore lost the race. He appeared pleased that the GEP would not let him get away with cheating to win. The boys had three more races, only one of which Peter won. The GEP asked Peter if he knew why he was losing and Peter said, "No." The GEP explained that Rex always tried harder. The boys had one more race and this time Peter really put a lot of effort into it and won by a long distance.

Almost immediately after trying hard and winning, Peter asked, "How is it you do that frog kick?" The GEP demonstrated and Peter practiced for a short while.

Peter also expressed some interest in learning the sidestroke and was shown how to do it. He caught on very quickly, but his rhythm was very poor. The GEP suggested that he practice the sidestroke. Peter swam the 20 feet across the pool about four times and quit. This swimming example is a good illustration of the fact that *for Peter to practice on almost anything he must obtain gratification through the practicing.*

This principle was explored and verified a week later when the boys practiced diving and swimming under water to retrieve objects off the bottom. This game was their own idea and, although they had considerable difficulty in finding the objects on the bottom of the pool, they stayed at it for some time. When the boys grew tired of this game, the GEP suggested a game of tag in which no one was permitted to get out of the pool. The boys agreed that this would be fun and played at it for about 30 minutes. During the game of tag, each boy had to swim

his best and Peter constantly was forced to put forth some effort to keep from being tagged.

### 3. *Tying Fishing Flies and Making Fishing Equipment as a Technique to Test Achievement Motivation and Obstacles*

A pre-existing interest of Peter's in fishing was utilized to stimulate interest in learning how to tie fishing flies and, through this, to learn more about the obstacles that interfered with achievement. Initially, he had to be given a great deal of help, so that the finished product was a nice looking fly; had Peter not succeeded on the first attempt he probably would have refused to make any more flies. Gradually assistance was removed (as Peter continued making flies during the following weeks) until he not only learned to make a fly on his own, but also became aware of how much help he had received initially. This was one of the first activities in which Peter learned to sustain effort at a task in which initially he had no skill and to maintain the effort until he became skillful.

The GEP asked Peter and Rex if they would like to make some fishing flies. Peter remarked, "It's awful difficult though, isn't it?" When the GEP said that it was not difficult, both boys were surprised and eager to try it.

While Peter tied the first fly, the GEP sat beside him—giving instruction and help with each step—so that Peter on his first attempt tied a nice-looking fly. He was very pleased.

While his younger brother tied a fly, Peter decided to tie another type of fly called a water bug, which is constructed with pieces of rubber. The GEP showed Peter how to measure and cut the rubber for the body of the water bug, by approximating the length of the hook's shank. A demonstration was given on how to thread the rubber body on the hook. At first Peter tried threading the body on the hook in the prescribed manner, but had considerable difficulty with this approach. He tore up two pieces of rubber and still had not succeeded.

While the GEP was busy helping Rex, Peter developed a better technique on his own. He unfolded a paper clip and poked a hole through the middle of the rubber body. Then he forced the eye of the hook through the hole he had made in the rubber body and it slipped on very easily. The technique that Peter developed was more effective and quicker than the one he had been shown by the GEP. Peter asked about the next step and was shown this. When the fly was completed, clear fingernail polish was applied to the thread used to divide the body into segments. Peter was not content to let the material sit and dry before proceeding on to the next step. Instead, he blew on it and shook the fly to make it dry quicker.

By this time it was clear that Peter's dexterity was extremely good and that he was resourceful. He caught on quickly to fly tying. He had

no difficulty making any of the small motor movements involved and made nice looking flies.

After Rex finished his fly, Peter began another one. He selected the color of materials he wanted to use and, as he began tying his fly, the GEP noted that Peter had forgotten much of what he had been told earlier and that certain points needed a lot of re-emphasis. Again the GEP sat beside Peter giving him the help he needed.

Peter continually referred to how many flies Rex had made in comparison to what he had done. He emphasized the fact that Rex had made more and better flies and had worked faster. He did not say anything derogatory about Rex's flies, but evidently *rivalry feelings* were developing.

Possibly this may have contributed to increased trouble soon to develop. As Peter tied another fly, he had considerable difficulty with the hackle (feather); when he wrapped the feather around the body of the fly and secured it with thread, he did not keep enough tension on the thread. Therefore, after he cut off the excess feather and attempted to throw some slipknots over the head of the fly, the feather unraveled since the feather had been cut off at the correct length and it was very difficult to fix the fly. Peter turned to the GEP and asked him to fix it. He conveyed extreme faith in the GEP's ability to fix his fly, so that his efforts would not have been wasted. The GEP did manage to straighten out Peter's fly. Although Peter had needed help with the fly and had asked for it, when the fly was completed he regarded it as a product of his labors alone.

The following week, when Peter and Rex were asked what they would like to do, both boys said that they wanted to tie more flies. The GEP stated that it would be necessary first to sort out the good feathers from the defective ones. Both boys agreed to help with this task but, during the process of sorting, *the boys competed with each other, which increased the time required to sort the feathers.*

The next week Peter asked if he might take some fly-tying materials home in order to make a few flies. The GEP stated that he could, just as soon as he had a vise. The GEP suggested that Peter try to make one. During the intervening week between sessions, Peter had gone out to his garage and looked for something out of which to make a vise; but he stated he could not find anything. The GEP told Peter that a vise could be made out of any one of a dozen things and then showed Peter a simple inexpensive vise made of wood. Peter was surprised that a vise could be made so easily.

The GEP then told Peter that fishing flies, like those he was making, sold for 35 to 55 cents apiece and cost only about one and one-half cents to make. Also, the GEP mentioned that an owner of a local sporting-goods store had said he would buy their flies if they made good ones. These comments seemed to spark Peter's interest a little more, and he stated that he would like to make flies to sell.

This week, when Peter and Rex tied their flies, they were not given

the assistance they had received in the past. Rex had forgotten a few steps, while Peter remembered all the steps necessary in constructing a fly. Peter, however, was making poor flies at first. He did not cover enough of the hook with the fly-body material, nor did he make even wings with the feathers. After making two or three flies, Peter became a little disgusted with himself because the flies he was tying were not as good as those he had tied initially. He did not realize that he was *now making flies on his own*, whereas previously he had received a lot of help.

Several times Peter commented that Rex's flies were better than his, which was true, but mainly only on those flies on which Rex had received some assistance. As the boys continued making flies, they would compare their flies and Peter would remark that Rex's flies were better than his. Each fly was evaluated more critically today, and the GEP gave pointers on improving them. Peter was shown where his flies were defective and on the next fly he sought to improve and, for the most part, was successful. Although a competitive attitude continued, *Peter did sustain his effort toward a goal.*

The fly tying was interrupted for an hour in order to swim at the YMCA, after which Peter and Rex returned to the flies. After this rest, Peter improved. Also, he could *straighten out his own mistakes* and he began correcting mistakes he had made on previously tied flies. In addition, he began to get the idea of how to make his fly look more like an insect.

Peter's *ingenuity continued to emerge.* The next week, soon after Peter climbed into the GEP's car to go home, he (Peter) stated that he had been thinking of a way of making a fly-tying vise. He thought he would take a couple of pieces of metal and put a bolt through the bottom and one through the top which could then be tightened to hold the fish hook. Peter also mentioned that in the garage he had found some metal that he could probably use for this part, but he had not figured out how to make a stand for the vise. The GEP told Peter his idea was a good one. Together, Peter and the GEP worked on an idea for the stand of the vise and then proceeded to construct out of scrap metal what they had designed.

The following week, both Peter and Rex showed the GEP flies that they had tied at home. Peter pointed out the defects in some of his flies and told how he *planned to make them better.*

As previously mentioned, Peter often expressed an interest in fishing, but we did not have any fishing equipment. We then sought ways of obtaining fishing equipment at the least possible expense, so that we could help Peter discover that much could be accomplished without spending a lot of money. Obviously, excellent rods and reels could have been purchased for Peter, but for two reasons this was not done. First, it had become apparent that Peter already had a problem of expecting to be given to, as his mother had given everything to her favorite baby. Second, one of the purposes of the growth



experiences was to find out whether Peter could respond to the idea of locating resources within the means of his family. New rods and reels would be out of the question. Therefore, we decided to try to make our fishing rods. A local junkyard was visited, and there broken radio antennas were obtained for 10 cents apiece. The man at the junkyard had told us (after learning what we had planned to do with the antennas) that it would never work; whereupon Peter became *discouraged*, as we had seen him get discouraged before. Peter was then told that some people have trouble accepting new ideas and that the GEP felt it would be worthwhile to try to make fishing rods from the antennas. We found that Peter could respond to this support; his attitude changed and he began talking about famous inventors whom the general public had tried to discourage by saying, "It won't work."

#### D. NEW LIGHT ON PROBLEMS AT SCHOOL

The achievement just discussed showed us another of Peter's difficulties at school. The making of the fishing flies was one of the first events that Peter chose to tell his sixth-grade fellow classmates about during a "Show and Tell" period, probably because his family could provide so little for him to bring to this class. After the growth experiences began, he remained shy and hesitant in confronting the class. At first, he took things that he made or learned to the teacher and asked her to show them to the class. Then he gradually began adding comments to the teacher's statements about the objects he brought to class and later *became able to present by himself the things he had brought*. As Peter became more involved in the "Show and Tell" period, his old hostile attitude toward school began breaking down and he began to look forward to telling the class how to make fishing flies, rods, and reels. Now that he had been given to, and had something to give, he felt more adequate to the opportunities for class participation.

#### E. "GOING FISHING" AS AN EXPLORATION OF LEARNING MOTIVATION IN AN INTEREST AREA

Peter had gone fishing many times with his father and brothers, but they had never caught many fish. It soon became apparent that Peter's ignorance of the habits of the fish resulted in his catching very few fish and this led to his defensive scoffing at fishing in the local lake. However, gradually he did learn through experience that even fishing is something to be thought about if you want to come home with a full string of fish. He learned where the fish were at different times of the year and the preferences of different species of fishes for different depths of water and varieties of food. He also learned



how to vary his angling techniques so as to hook the type of fish he was after. Peter was not always an eager pupil but, after his first trip home when he was struggling up to the house under the weight of more than 100 crappie, his attitude toward fishing in the local lake and learning the habits of fish began to change. Peter began to see the difference between the way his father fished and the way we fished and he took great *pride in carrying home his new-found knowledge, his first step toward using his own ability even if it involved doing something better than his father had done it,*

The first time Peter and Rex were taken fishing, the GEP discovered that they knew next to nothing about this activity. As the boys and the GEP drove out to the local lake, the GEP stated that he thought the fishing would be pretty good. Peter wanted to know why he thought so; so the GEP explained that it was a cloudy day and looked like it would rain. He went on to state that it had not rained for a while and that fishing is not very good immediately after a heavy rain because the rain knocks insects into the water, as well as washing insects, worms, etc., off the banks into the water enabling the fish to feed, so that they are no longer hungry. He also explained that just before a rain and during a rain was a good time to fish because then the fish were looking for food.

The GEP then explained that they would try a cove at the south end of the lake, and again Peter wanted to know why. The GEP explained that during the early spring the fish were in the shallower water which was warmer, and that as it became still warmer the fish sought the deeper water which was cooler. Peter was then told that the shallower water was at the south end of the lake; that was why we would fish there.

The boys were fishing for crappie and bluegill, but neither boy had any idea of how to string his pole correctly. They tended to use hooks so large that the fish could not get them into their mouths and to put on such heavy weights that their bobbers would sink. In addition, the only bait they planned to use was worms, which they did not know how to put on correctly. After they were shown how to fix their poles, they had to be taught how to cast.

Peter caught on to casting very quickly and did an excellent job of it. The boys did not have any luck fishing with worms. The GEP told them that, when fishing for crappie, minnows were much better bait. Later a fisherman gave the boys some minnows. Like most boys, Peter wanted to hook the minnow through the mouth, but was shown how hooking the minnow just behind the dorsal fin allows the minnow to swim relatively freely and thus attract more fish. The boys did catch a few crappie. Besides the few fish the boys caught, they accepted from several fishermen all the fish the latter had not wanted, so that the boys ended up with over 100 fish. Peter was very pleased to be able to provide his family with so much food, but had never cleaned any fish.

At the boys' home, Peter got some papers, paring knives, and a pan

of water. He was then shown how to scale the fish, cut off the fins, and remove the entrails.

In this sequence of experiences, it became obvious again that Peter was very receptive and *quick at learning something necessary to the pursuit of a real interest of his own.*

A couple of weeks later, Peter continually laid his fishing rod on the dock and expected people not to step on it. Continually, he had to be reminded to set it out of the way. Again, both boys had to be shown how to string their poles. The minnows they used were a little large and the GEP repeatedly told them to wait a little longer than usual for the fish to get such a big minnow in its mouth. The boys just kept losing their minnows and catching very few fish. Later, they were given some small minnows which enabled them to catch several fish. Peter and Rex were having so much fun that when it was time to leave they ignored the GEP's statements that it was time to go. Finally, after the GEP reminded them that we could come back next week, they prepared to leave.

The following week the boys fished at another lake, mainly for bluegill. A fish grabbed Rex's hook, took off, and broke the line. The bobber skipped along the water as the fish headed for the middle of the lake. Peter started into the lake after the bobber and fish, but was stopped by the GEP. Peter was really disappointed to see that "big one" get away.

In the fishing situation, Peter's rivalry was constructively used. Rex made a good cast and got his line out quite far. Peter then tried to cast his line out that far, but failed on his first attempt. The GEP suggested that he leave it where it was, but Peter said, "No, I can't let my little brother beat me." He continued casting until he managed to get his line out a little farther than Rex's.

A week later, both Peter and Rex were jerking too hard when they got a bite. Peter jerked so hard that his bobber slammed against the rocks and broke. He put another bobber on and did the same thing. After that he decided that he had better fish without a bobber, which he did for the rest of the afternoon.

Peter now not only catches on more quickly than Rex, but is usually more careful as well. However, he gives Rex a real rough time when he (Rex) does something incorrectly or has an accident. For example, when Rex accidentally knocked over the minnow bucket and spilled some of the minnows into the lake, Peter said, "Oh, damn you, Rex." Peter also becomes quite angry when he tries to show Rex how to cast better, and Rex does not catch on quickly and often says, "Rex, you're a fool" or "You're stupid." In other words, while the growth experiences showed Peter's capacity to channel aggression in new directions, it also showed how little control or modulation of anger he had at the time.

Peter grew tired of catching bluegill and had the GEP help him set up his rod and fish for catfish. He did not watch his line; so he never knew when he had a bite. When Peter caught a bluegill (while he was

fishing for catfish) he would say, "Oh, damn," and throw it back. Previously he had been keeping fish of the same size.

A couple of weeks later Peter wanted to fish for bullheads again and was annoyed that Rex preferred to fish for bluegill. Peter tried several times to cast his line way out, but failed. He then asked the GEP to do it for him. Peter was told (after his line was out) that he would have to sit and wait. Peter did not like this; he impatiently wanted to reel in his line. After the GEP followed his own suggestions and caught a nice-size bullhead, Peter was much more willing to sit quietly and watch his line rather than play with it all the time. Peter caught a small bullhead, but later Rex caught two good-size bullheads at once. Peter was astonished; he could not believe it, but neither could he deny it. Later the GEP also caught two at once, and although Peter had been annoyed and angry when Rex caught two at once, he seemed pleased and derived some vicarious satisfaction from the GEP catching two at once. Later Peter caught two fish at once and was thrilled.

Sometimes when casting, Peter or Rex would cast his line over the other one's line and a few angry words flew back and forth. Peter would call Rex a dumb fool and Rex would angrily tell Peter it was his own fault.

A couple of months later: Peter had been told many times how some fish will stay at a particular depth in the water and that when you see someone near you catching fish, to note how deep he is fishing. In the past, Peter had not only not paid any attention to the depth, but when he was told how deep to fish he often would not pay any attention. Today, however, Peter noted how deep the successful fisherman's line was set and would fish at that depth. He caught several fish.

Several times, when Peter was putting fish in the net, they would fin him and he would drop them into the water. When this happened he would look sick and very unhappy with himself. He had been told many times to put his finger in the fish's mouth or gills, but only after he dropped several fish back into the lake when they finned him did he begin to follow the GEP's suggestion.

He seemed at times to have to learn from experience.

One day we rented a boat and, when it began to rain, we had tied it to a covered dock. After the rain slowed down, Peter decided to fish from the boat. He stood on a seat of the boat, although the GEP had told him not to do so, as he might fall into the lake. It wasn't long before Peter lost his balance and fell into the lake. He was soaked and, as it had become colder, the fishing trip ended. Peter never stood up in a boat again.

Peter sometimes talked about a person, a fishing rod, or a particular fishing spot as being jinxed. On one occasion, Peter stated that the GEP was jinxed because he was not catching any fish. The GEP explained to Peter that the wind was strong and that his light fly rod was moving with each gust of the wind, so that he could not feel the light bites of the crappie. Peter was disbelieving, so the GEP borrowed

Peter's rod and quickly caught three fish. Peter then decided that the GEP was not jinxed, but rather that his rod was jinxed. The GEP explained to Peter that a term like "jinxed" was used only by people who could not understand the reason for their lack of success, and that he (Peter) did not need to use such a term because he knew why the GEP had not been catching fish.

#### F. REVIEW OF EVIDENCE OF PETER'S CAPACITY FOR CHANGE

During July of 1961, five months after the original interview with Peter's teacher, the same teacher was again interviewed. The teacher reported that Peter had not engaged in a fight for more than a month, that he was participating in activities when he knew he could not be first or best, that he had developed quite an interest in reading, and that he got along well with the other students. On the negative side, the teacher stated that Peter had trouble with spelling and that he *would not even try to learn*. He would also refuse to be a class monitor, apparently because he was afraid of making a mistake. Also, he still dealt with frustration by *trying to change the rules in games to suit his own rules*, but the other students would not let him get away with it.

His marks at the end of the sixth grade, however, still indicated that he was not using much of his ability. He had an F in spelling, a D in writing, and all of his other grades were 'Cs—except for arithmetic and reading in which subjects he earned Bs.

#### G. EVIDENCE OF EMERGENCE OF RESPONSIVENESS TO WORK CHALLENGE

During the last interview with Peter's sixth-grade teacher, she had reported that Peter had become increasingly aware of his family's limited finances. One of his major concerns was his old, worn, patched clothing. This suggested that it might be worthwhile to explore some work experiences for Peter to see whether he could be motivated to earn money for new clothes the next fall. It was also thought it would be worthwhile (a) to find out *whether he could learn to work for what he wanted* and (b) to *discover the nature of conflicts which were contributing to work difficulties*.

During the summer of 1961, Peter and Rex spent two mornings a week improving a wooded picnic area, for which work they were paid 50 cents an hour. The work included raking and hauling leaves, planting grass, quarrying and carrying rock, and building a rock retaining wall. Peter did not work well at certain tasks, such as raking leaves or carrying rock. While engaged in this repetitious type of work, he tended to rest a great deal and to work very slowly. An effort was then made to see whether he would



respond to monetary reward and punishment. The situation was handled by reducing Peter's pay by one-half whenever he worked at half pace and by making him aware of the fact that most employers demand a full day's work for a full day's pay. Peter responded to these techniques, but still did not work steadily without a great deal of supervision. In other words, we had to infer that *money alone was not a dependable motive to evoke sustained effort.*

We then compared the tasks on which he worked energetically with those on which he dawdled and observed that, when Peter was engaged in more strenuous and particularly *challenging work* new to him (such as quarrying rock, laying a rock wall, making a dry wall, or staining a picnic table), he worked with a great deal of persistence, enthusiasm, and without much supervision. On most of these tasks he worked well alone. All of these tasks were "a man's work," not tasks which could be done easily by a child, as was leaf-raking. This suggested to him the importance of respect for the job, of *wanting to do something consistent with his rapid growth in size*—which meant that he was getting to be a man.

#### H. TESTING POTENTIAL LONG-TIME MOTIVATION

Due to Peter's difficulty in sustaining effort at a repetitious task, an effort was made to explore ways in which long-term motivation could be developed. One long-term project, the construction of a pontoon boat, was selected by Peter from several projects which were suggested. Building a boat would test whether he could maintain the necessary interest and industry required to finish a relatively interesting long-term project, but one which still involved considerable repetitious effort.

Because of the cost of wood and other supplies, it was explained to Peter that it was not practical to start immediately as that procedure would involve paying more for supplies than we were able to spend. But, in the spring, several old garages at the Foundation were torn down, and we were given permission to take the wood we needed for the construction of a pontoon boat.

Peter was enthusiastic about building the boat and, from the start, he worked more like a man than a boy. The sections of the old garage were torn apart, nails were removed, and the boards were run through an electric planer. Rough plans of the boat were drawn by the GEP with Peter's assistance as the materials were ready for construction. With only a handsaw, Peter cut all the necessary two-by-fours for the frame of the boat. He also cut the siding used for the floor of the craft. While this may not seem to be



much work, it should be realized that the pontoon boat is 12 feet long and eight feet wide, and that Peter (with very few breaks) frequently did nothing but saw boards for four hours. On a challenging task, then, he could maintain his effort for sustained periods of time.

While taking Peter and Rex to and from the Foundation, we discussed other materials we needed. Due to the cost of pontoons, we decided to use eight 55-gallon barrels, four to each side, which Peter took the initiative of securing (through his father's place of employment) for \$1.50 per barrel. As the work progressed, Peter's interest increased, as did the effort that he put into the work. His only "pay" consisted of a few pieces of candy, the number of pieces varying according to his effort. As the craft neared completion, Peter even asked to work extra time in the evening so as to hurry it along. This he did without loss of interest or reduction of effort. The construction was not without its setbacks. Once, the entire frame had to be torn apart and put back together. Twice, the cables which held the barrels secure to the frame had to be removed and replaced. Of course, Peter was sorry to see these setbacks, but he took them quite well because this was to be "our boat" and we wanted it to be well made.

#### I. EVIDENCE OF CAPACITY FOR CHANGE IN CERTAIN AREAS

In the fall of 1961, Peter entered junior high at a new school. He expressed a desire to go out for football and was encouraged to participate in athletics and to cooperate with the coach. This time he stayed out for football, identified strongly with the school, and pointed out with pride their star fullback, a colored boy. Peter also participated in wrestling, in which sport he came close to earning a letter. In the spring, he was taught to play tennis and tried to go out for the tennis team, but unfortunately the season was already half over.

During the spring of 1962, Peter's seventh-grade teachers were interviewed. The general consensus of the teachers' remarks ran as follows: On the positive side, Peter does well in tests in a couple of his classes; his oral work is better than his written work; he is well behaved and well liked; he has a lot of ability which he does not use; and his knowledge of subjects not directly related to school surpasses that of most of his classmates. But he still wastes more time than the average student; he usually brings electronic gadgets to play with; he usually does not follow instructions well; he writes illegibly; he rushes to finish his work and seldom checks his work; *if he does not see any use for something, he refuses to study it*; his friends are not good students and some of them have been in trouble—stealing.

We see further evidence here of some of our conclusions: Peter would not focus on tasks (including listening to instructions, carrying out the task) meaningless to him; he expressed his wish that school should deal with real or challenging tasks by bringing along his electronic gadgets. He evidently depreciates himself, judging by the company he keeps.

At the end of the fifth six-week period, Peter had received two Bs, two Cs, and two Ds. After his interview with Peter's teachers, the GEP talked with Peter about school. *Peter saw no use for most of his school work.* Its usefulness for any occupation was explained by the GEP to Peter. When school ended, Peter, beaming, brought his grade card to the GEP. In three subjects, he had maintained his grades; while in the other three, all marks were raised one grade. Peter was praised generously for this achievement.

Although Peter's academic achievements were still far from his non-academic achievement and his capacity, he had made some improvements. Also, in contrast with his previous lack of participation in organized athletics, he had been out for two athletic teams all season and had made efforts to engage in a third. Of no less importance is the fact that Peter no longer hated Negroes; he was proud to have them on the school athletic teams. Evidently, *as his own achievements in and out of school improved, he no longer needed a scapegoat or outlet for his anger or frustration.*

#### J. EVIDENCE OF CAPACITY TO COMMUNICATE AND BUILD UPON HIS OWN INTERESTS

Increasingly, Peter communicated an interest of his own either by mentioning it or by bringing it along, as the following example illustrates.

Peter brought along a lizard on one growth experience session and asked, "What kind of lizard is it? What does it eat?" and "What should I keep it in?" Peter and the GEP then visited the city library, where they established the genus and species of the lizard. At the same time, the purpose of classifying animals and reptiles was explained to Peter. A local biological supply house was then visited and Peter learned that meal worms are used to feed lizards, as well as at what stage in their development meal worms are useful food for reptiles.

Peter was told that an aquarium would be a suitable place to keep the lizard. A tour was then made of local secondhand stores, in one of which an old five-gallon aquarium was obtained for \$1.00. The aquarium was purchased by the GEP for Peter to use with the understanding that he could keep the aquarium as long as he kept something alive in it. Here again would be a test of his capacity to sustain effort. A periodic check was made to see

if Peter was fulfilling his end of the bargain. When the lizard choked to death on a piece of jelly which Peter's younger brother had given it, Peter bought two goldfish for the aquarium. This involved more learning about the management of an aquarium. He was then taught to wash sand for the bottom of the aquarium so as not to discolor the water, and then vegetation was planted in the sand. With the introduction of plants into the aquarium, information about photosynthesis (which Peter had obtained in grade school) was put to use, and was supplemented with more detailed information. Approximately one year after the introduction of the fish into the aquarium, one of Peter's younger sisters accidentally pushed the doorknob through the front glass of the aquarium. Peter was first quite angry with his sister, but then (as he learned to repair the aquarium) became so interested in how aquariums are put together and sealed that his anger subsided and his interest and enthusiasm increased. As Peter repaired the aquarium, he frequently talked of the interesting and colorful tropical fish which he had seen in an aquarium owned by an elderly lady whose lawn Peter had frequently mowed. It was explained to Peter that he could purchase the aquarium he had been using and that, if he liked tropical fish, there was no reason why he could not use some of the money he had made mowing lawns to purchase some fish and a secondhand aquarium heater, an air pump, and a thermometer. Peter was jubilant.

Peter and the GEP spent several weeks watching the want ads in the local newspapers and checking with secondhand stores until a good used aquarium heater and an air pump were located and purchased. In this case, as in many others, the items were purchased by the GEP; then Peter paid small weekly amounts until the purchase price was paid. Next, Peter and the GEP designed and constructed an aquarium filter to eliminate the necessity of cleaning the aquarium. Having obtained the necessary equipment for raising tropical fish, Peter was then taken to an aquarium store where he purchased his first tropical fish.

A short time after Peter purchased his fish, several young fish were born and Peter was as thrilled as a new father. He asked about breeding tropical fish and his questions were answered from a paperback book on the breeding of fish, which book was loaned to him. He read the book from cover to cover several times. A few weeks later his female fish had a second batch of young. As Peter's fish grew, so did Peter's interest in raising them until, at present, he has three aquariums and all the necessary equipment to go with them, which he has paid for himself.

Here we see in vivid contrast Peter's indifference to intellectual work,

which seemed meaningless to him, and his eagerness to get information (even from books) when it was needed for an activity in which he had a deep investment.

#### K. SUMMARY

We hope that it is clear to the reader how the use of these Growth Experiences revealed attitudes, feelings, conflicts, obstacles, potentialities, and strengths in Peter which standard clinical examinations had left unclear. We also believe that these experiences revealed Peter's responsiveness to a wide range of opportunities directly related to his drives, developmental problems, and interests—thus adding to our understanding of Peter as well as opening new areas of interest for him. In addition, we saw how certain activities (sling, fishing, swimming play) provided for redirecting aggressive drives; these also provided gratification enough, so that previously hostile and destructive reactions to frustration subsided. These growth-experience activities, then, tailored to individual needs as progressively discovered, provided a tool that led to significant answers to some of our questions regarding the developmental disturbances in this potentially able or even gifted boy. A full report on the dynamics of his development will be published at a later date.

*The Menninger Foundation*  
*Topeka, Kansas 66601*



## EGO IDENTITY AND SOME ASPECTS OF PERSONAL ADJUSTMENT\*<sup>1</sup>

*Department of Psychology, Fordham University*

---

S. M. HOWARD AND J. F. KUBIS

---

### A. INTRODUCTION

Erikson's recent contributions (5, 6) to ego psychology stress the conflicts that occur within the ego rather than those occurring among the psychic structures. Two emphases characterize his approach: a concern with the continuity of the self and a description of growth in terms of an expanding social radius. Under the pressure of a rapid expansion of social contacts, the adolescent continually scrutinizes his own sense of sameness and continuity of self. This being so, the present research deals with the sense of self-continuity, called *ego identity* by Erikson, and its relationship to personal adjustment in college women.

Erikson (7) has presented a theory of ego development that postulates eight stages within the life-span, each of which embodies an intrinsic conflict to be resolved. Identity *vs.* role diffusion is the conflict prominent during the fifth stage, adolescence. Erikson refers to this period as the "identity crisis," when the adolescent, confronted with long-term commitments, redefines himself in the light of new goals. At a time of rapid physical growth and genital maturity, the adolescent begins to question his sameness and the continuity of his past (6). Further, as his contacts with society widen, the adolescent must match his present skills and roles to those of social and occupational prototypes of the times. In effect, he must synthesize his past, present, and future.

Within the framework of this theory, identity is considered as a complex of self-referent images which, gradually evolving through social interactions, produces a delineation of the self. A useful distinction was made by Bronson (3) between the content and the relationship among the component self-images. The latter dimension he called *identity structure*,<sup>2</sup> the pattern of self-referent ideation. Further, he presents evidence for two dimensions of identity structure: diffusion of images and isolation of images.

---

\* Received in the Editorial Office on July 14, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This report is based on a study submitted by the first author as a Doctoral dissertation to the Graduate Faculty of Fordham University, May, 1963.



Because ego identity reflects the integration among the component self-referent images, congruence of self-images within the matrix means less inner tension, less anxiety. Bronson (3) found, in fact, that ratings on implicit anxiety correlate significantly with his two dimensions of identity structure. He interpreted anxiety as a correlate of poor integration within the identity structure.

Several other studies have investigated various aspects of personal adjustment and ego identity. Gruen (9) found that individuals, unsure of their identity and role, manifest instability of self-evaluation. Howard (10), studying high-school students, reported that girls with greater identity diffusion feel more isolated from one or both parents. In a study on identity diffusion by Block (2), interpersonal consistency (as a correlate of ego identity) showed a positive relationship to adjustment. Within a sample of navy recruits, Rassmussen (13) found that poorly adjusted men and those with character disorders have less well-defined ego identity than have the best adjusted men. Consistent with these findings are those of Thompson (17), who obtained positive correlations between ego identity and personality traits indicative of good adjustment.

No studies have directly related anxiety to ego identity. Howard's research (11) has shown the importance of maternal identification for ego-identity formation in college women. A faulty mother-daughter relationship may elicit a number of significant reactions. First, poor mother-daughter relations may create basic anxiety within the personality. Secondly, this interaction may be the source of the daughter's hostility and, in turn, of further anxiety about hostile attitudes. Finally, this unsatisfactory interaction may generate between mother and daughter a number of problems, both implicit and explicit. The purpose of the present study was to relate ego identity to these three fundamental aspects of personal adjustment: manifest anxiety, manifest hostility, and mother-daughter problems. The hypothesis to be investigated is that ego identity in college women is negatively related to the three foregoing measures of maladjustment.

## B. PROCEDURE

### 1. *Subjects*

The sample consisted of 130 freshmen and 115 sophomores, American born, ranging in age from 17 to 20 years. Their parents were living together and all were of middle-class socioeconomic level. All students resided on the campus of a Catholic women's liberal-arts college in the Midwest.

## 2. Test Instruments

Five test instruments were administered: an Ego-Identity Scale, a Manifest-Anxiety Scale, a Manifest-Hostility Scale, a Mother-Daughter Problem Checklist, and a Rating Scale for Identity Traits.

a. *Ego-Identity Scale*. The Ego-Identity Scale (11) consists of 50 items indicative of various personality dimensions relevant to ego identity. Odd-even reliability coefficients for 130 freshmen and 115 sophomores were .74 and .64 respectively, and test-retest coefficients for a one-week interval were .72 and .78 for a smaller group of 83 freshmen and 96 sophomores. From the smaller group of subjects, an estimate of the validity of the Ego-Identity Scale was obtained by correlating ego-identity scores with self-ratings on the Rating Scale for Identity Traits. The freshman coefficient was .34; the sophomore coefficient, .60.

b. *Manifest-Anxiety Scale (MAS)*. This scale, derived by Taylor (15) from the Minnesota Multiphasic Personality Inventory (MMPI), consists of 50 items indicative of manifest anxiety. In the present study, the 15 items which Taylor included from the Lie Scale were also used.

c. *Manifest-Hostility Scale (MHS)*. Like the Taylor Scale, the Manifest-Hostility Scale (14) contains 50 items. These items were randomly combined with those of the MAS. When the MAS and the MHS were combined, Feldman and Siegel (8) found no significant changes in the MAS and MHS scores.

d. *Mother-Daughter Problem Checklist*. This instrument (4) consists of a list of 60 problems involving interpersonal relations and attitudes. Some items were adapted from the three levels of the Mooney Problem Check List (12); others, from an unpublished biographical inventory for adolescent girls and from areas of mother-daughter relations cited by Thomas (16). The present authors' experience with girls of college age suggested the remainder of the items.

Fifty-five items submitted to a clinical psychologist were later revised to decrease ambiguity. Other items were added, and 67 items were grouped into four categories: Presumed Deficiencies of the Mother, Rapport with the Mother, Maternal Domination, and Maternal Rejection. For an appraisal of content validity, the revised list was sent to five experienced counselors of college girls. After evaluation of the items, 15 statements were selected for each category. The number of problems checked was tallied for Problem Count. For 130 freshmen and 115 sophomores, the odd-even reliability coefficients, corrected by the Spearman-Brown formula, were .89 and .92 respectively.

*e. Rating Scale for Identity Traits.* This scale was also devised for a previous study (4) to provide a criterion of personality traits indicative of identity. The summed ratings on each six-point scale for 17 paired traits constituted the score for Identity Traits.

### 3. Administration

To encourage candid responses, a procedure was utilized to assure anonymity for subjects. Two test sessions were held. During the first session, the Ego-Identity Scale, the Manifest-Anxiety Scale, the Manifest-Hostility Scale, and the Mother-Daughter Problem Checklist were administered. A week later, the Ego-Identity Scale was readministered, and students also rated themselves on the Rating Scale for Identity Traits.

## C. RESULTS

The means and standard deviations for the scores on ego identity, manifest anxiety, manifest hostility, and mother-daughter problems are shown in Table 1. The sophomores obtained higher ego-identity scores than did freshmen, a difference significant beyond the .01 level. Sophomores also earned

TABLE 1  
MEANS AND STANDARD DEVIATIONS FOR EGO IDENTITY, MANIFEST ANXIETY,  
MANIFEST HOSTILITY, AND PROBLEM COUNT

Test variables	Mean		Standard deviation	
	Freshman ( <i>N</i> = 130)	Sophomore ( <i>N</i> = 115)	Freshman ( <i>N</i> = 130)	Sophomore ( <i>N</i> = 115)
Ego Identity	34.58	37.42	5.55	4.73
Manifest Anxiety	18.80	16.23	8.04	7.62
Manifest Hostility	14.98	11.61	6.23	4.56
Problem Count <sup>a</sup>	25.05	24.86	11.13	10.50

<sup>a</sup> Transformed scores.

lower scores (indicating better adjustment) on the MAS, MHS, and Problem Checklist. The *t* ratios for class differences in anxiety were significant at the .01 level; those for hostility, at the .001 level. As for mother-daughter problem count, the small difference between the freshman and sophomore classes did not prove significant.

Table 2 presents the correlations of the three adjustment variables with ego-identity scores and identity-trait ratings. As expected, all coefficients of these measures of adjustment with the two indices of ego identity are negative. However, not all the correlations, as judged by a one-tailed test, are significant. The basic finding is the relatively high, consistent, and significant negative correlation between ego identity and anxiety.

TABLE 2  
CORRELATIONS OF EGO-IDENTITY SCORES AND IDENTITY-TRAIT RATINGS WITH  
THREE MEASURES OF ADJUSTMENT

Measures of adjustment	Ego-identity scores		Identity-trait ratings	
	Freshman ( <i>N</i> = 130)	Sophomore ( <i>N</i> = 115)	Freshman ( <i>N</i> = 83)	Sophomore ( <i>N</i> = 96)
Manifest Anxiety	-.48**	-.46**	-.33**	-.45**
Manifest Hostility	-.15*	-.07	-.06	-.06
Problem Count	-.13	-.17*	-.25*	-.15

\*Significant at the .05 level.

\*\* Significant at the .01 level.

From Table 2 it is evident that only the freshman correlation between ego identity and manifest hostility is significant (.05 level). On the whole, the correlations for ego identity and problem count are higher than are those for identity and hostility. Two of these coefficients are significant: at the .05 level, that between ego-identity scores and problem count (for sophomores) and, at the .01 level, that between identity-trait ratings and problem count (for freshmen).

Intercorrelations among the three measures of adjustment for the two classes are shown in Table 3. All the intercorrelations are significant at the

TABLE 3  
INTERCORRELATIONS AMONG THREE MEASURES OF ADJUSTMENT IN FRESHMAN  
AND SOPHOMORE COLLEGE WOMEN

Scale	Problem Count	Manifest Hostility
<i>Freshman (N = 130)</i>		
Manifest Anxiety	.16*	.48**
Manifest Hostility	.41**	
<i>Sophomore (N = 115)</i>		
Manifest Anxiety	.30**	.41**
Manifest Hostility	.43**	

\* Significant at the .05 level.

\*\* Significant at the .01 level.

.01 level except that between manifest anxiety and problem count for freshmen, which is significant at the .05 level. These results substantially confirm expectations.

#### D. DISCUSSION

Results indicate that freshmen and sophomores differ significantly on measures of identity and adjustment. Sophomores show higher ego-identity scores than do freshmen, and therefore they appear to demonstrate a more effective resolution of the "identity crisis" (11). Consistent with their lower ego-identity scores are the higher-anxiety and higher-hostility scores obtained by



freshmen. Testing out newly acquired roles, adjusting to recent emancipation from home and parents, and responding to increased academic pressures are anxiety-producing experiences. Results from the present investigation accord fairly well with those of Barta (1), who found greater anxiety in male college freshmen than in sophomores.

For both classes, as predicted, manifest anxiety shows a negative relationship to both measures of identity: the Ego-Identity Scale and the Rating Scale for Identity Traits. It was expected that a lack of integration among component self-images within the identity structure would result in anxiety. The findings indicate that the more cohesive a student's personality structure (that is, the better the ego identity), the less anxiety he manifests. These results are in agreement with those of Bronson (3), who found that implicit anxiety—rated in interviews—correlated positively with identity diffusion. Present findings are in accord, too, with those of Block (2), who observed a positive relationship between anxiety and lack of role variability.

Manifest hostility failed to show the same well-defined relationship to ego identity that obtained between anxiety and ego identity. Only one correlation, between freshman ego-identity scores and manifest hostility, is significant and it is low. It is not altogether clear why ego identity and manifest hostility are not so closely related as are ego identity and anxiety. One possible explanation may be that as anxiety mounts, ego identity weakens correspondingly, but only with a great deal of anxiety does manifest hostility appear. A second possibility may be that the items of the MHS are more visible than are those of the MAS and thus are less acceptable to students whose religious beliefs discourage hostile attitudes and behavior. While many items of the MAS treat of physical symptoms over which the student feels she has no control, a number of statements in the MHS concern behavior for which the student may feel responsible and therefore reluctant to admit.

Of the correlations between problem count and measures of identity, two of the four coefficients are significant: that between problem count and freshman self-ratings on identity traits, and that between problem count and sophomore ego-identity scores. The items of the Mother-Daughter Checklist were completely visible, so that a cultural factor may account for the lack of consistent results. In the Christian tradition, criticism of parents and authority figures is frowned upon and hence might preclude an open acknowledgment of difficulties with the mother.

While no other studies have related ego identity to hostility or mother-daughter relationships, several recent investigations confirm the positive relationship between ego identity and adjustment. Rassmussen's two groups of



navy recruits with poorer adjustment had lower ego-identity scores than had a well-adjusted group (13). The present findings of a negative relationship between mother-daughter problems and ego identity is in accord with Howard's results (10) showing that high-school girls with poorer identity feel a greater emotional distance from one or both parents. Because readily fluctuating self-images indicate poor adjustment, Gruen's report of instability of self-evaluation as a correlate of poor identity (9) presents additional evidence for a negative relationship between ego identity and unsatisfactory personal adjustment. Thompson (17) presents direct evidence for this relationship. She found significant correlations between six factors on Cattell's Sixteen Personality Factor Questionnaire (16 P.F.) and ego-identity scores. The "components of identity" cited by Thompson, which may be considered as leading to adequate adjustment, are ego strength, composure, self-confidence, social adventurousness, character, and superego.

Although the foregoing studies related ego identity or its correlates to various facets of adjustment, the present study investigated a basic relationship: that between ego identity and manifest anxiety. The inverse relationship found between them is consistent with the findings of Howard's earlier experiment (11): namely, that ego identity is positively related to maternal identification (an identification of fundamental importance, one that is probably closely linked to anxiety). Further support for the conjecture that maternal identification is important for ego-identity formation in women comes from the present findings that ego identity and mother-daughter problems show an inverse relationship. From the results of the present research, it is evident that the degree of ego identity achieved by young women in late adolescence is reflected in their personal adjustment, including satisfying interactions with their mothers.

#### E. SUMMARY

Erikson's theory of ego development postulates an "identity crisis" at adolescence, issuing either in stabilization of ego identity or in identity diffusion. Previous research has suggested that anxiety is positively correlated with identity diffusion. The present experiment tested the hypothesis that ego identity in college women is inversely related to three basic indicators of adjustment: manifest anxiety, manifest hostility, and mother-daughter problems. It was found that ego identity is inversely related to manifest hostility and to the incidence of mother-daughter problems. However, the significant results are not consistent for freshmen and sophomores nor for the two measures of identity. On the other hand, ego identity manifests a strong, consistent inverse relationship to anxiety.

## REFERENCES

1. BARTA, J. A. A study of the concurrence of anxiety and hostility. Unpublished Doctoral dissertation, Fordham University, New York, 1962.
2. BLOCK, J. Ego identity, role variability and adjustment. *J. Consult. Psychol.*, 1961, **25**, 392-397.
3. BRONSON, G. W. Identity diffusion in late adolescence. *J. Abn. & Soc. Psychol.*, 1959, **59**, 414-417.
4. DIGNAN, S. M. H. Ego identity, identification, and adjustment in college women. Unpublished Doctoral dissertation, Fordham University, New York, 1963.
5. ERIKSON, E. H. Ego development and historical change. In A. Freud, H. Hartman, & E. Kris (Eds.), *Psychoanalytic Study of the Child (Vol. II)*. New York: Internat. Univ. Press, 1947. Pp. 359-396.
6. ———. *Childhood and Society*. New York: Norton, 1950.
7. ———. The problem of ego identity. *J. Amer. Psychoanal. Assoc.*, 1956, **4**, 56-119.
8. FELDMAN, M. J., & SIEGEL, S. M. The effect on self description of combining anxiety and hostility items on a single scale. *J. Clin. Psychol.*, 1958, **14**, 74-77.
9. GRUEN, W. Rejection of false information about one's self as an indication of ego identity. *J. Consult. Psychol.*, 1960, **24**, 231-233.
10. HOWARD, L. Identity conflicts in adolescent girls. *Smith Coll. Stud. Soc. Wk.*, 1960, **31**, 1-21.
11. HOWARD, S. M. Ego identity and maternal identification. *J. Abn. & Soc. Psychol.*, in press.
12. MOONEY, R. L., & GORDON, L. U. Mooney Problem Check List (1950 rev.) New York: Psychological Corp., 1950.
13. RASMUSSEN, J. E. An experimental approach to the concept of ego identity as related to character disorders. Unpublished Doctoral dissertation, American University, Washington, D.C., 1961.
14. SIEGEL, S. M. The relationship of hostility to authoritarianism. *J. Abn. & Soc. Psychol.*, 1956, **52**, 368-372.
15. TAYLOR, J. A personality scale of manifest anxiety. *J. Abn. & Soc. Psychol.*, 1953, **48**, 285-290.
16. THOMAS, R. *Mother-Daughter Relationships and Social Behavior*. Washington, D.C.: Catholic Univ., 1955.
17. THOMPSON, S. M. St. G. Modifications in identity: A study of the socialization process during a sister formation program. Unpublished Doctoral dissertation, University of Chicago, 1963.

*Department of Psychology*  
*Fordham University*  
*Bronx, New York 10458*

## SET AND SHOCK-STRESS EFFECTS UPON ILLUSION PERCEPTION\*<sup>1</sup>

*Department of Psychology, Boston College*

---

DANIEL J. BAER

---

### A. PURPOSE

An understanding of the effects of psychological stress upon perceptual-motor performance is of great theoretical and practical importance (1, 8). Although severe stress tends to inhibit such performance (3, 4, 6, 8), mild stress may increase efficiency (12). Several investigators (3, 5, 7) have suggested that changes in the perception of visual illusions are indicative of stress effects. Since mental set is also an important determiner of perceptual performance (2, 9),<sup>2</sup> its influence should interact with stress effects in such a manner that illusion perception will be characteristically modified.

The purpose of the present study was to evaluate the interactive effects of electric-shock stress and mental set upon the perception of visual illusions. It was expected that with mild stress and a facilitating mental set, illusion perception would be maximal, while with severe stress and an inhibiting mental set, illusion perception would be minimal. Under other set-stress conditions, it was expected that illusion perception would be within the limits of these two extreme conditions.

### B. METHOD

#### 1. *Subjects*

Ss were 63 male college students. Only those paid volunteers who reported an absence of neurological or sensory-motor difficulty were selected as Ss.

#### 2. *Apparatus and Stimuli*

The *S* was seated against one wall of a 7' by 7' cubical with his head positioned in a Bauch Lamb head-chin rest, and with the test stimuli mounted or displayed at eye level on the opposite wall. The equipment controlling the stimuli was outside the test room. An intercom permitted verbal communication between *E* and *S*, and a signal button enabled the *S* to activate a Veeder

---

\* Received in the Editorial Office on July 21, 1964, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> These data were gathered while the author was a postdoctoral research associate with the Training Research Branch of the Aerospace Medical Research Laboratories, Wright Patterson Air Force Base, Ohio.

counter and a .01-second clock. The duration of the stimuli was controlled by Hunter decade interval timers. Except for the test stimuli, no other light source was present in the cubical.

The *Ss* were shocked through finger electrodes attached to the index and small finger of the nondominant hand. Intensity of the shock was controlled by an Applegate Constant-Current Electric Stimulator and the shock duration by a Hunter Decade Interval Timer.

The illusions were presented in the following order:

a. *Necker cube*. The stimulus was a white-on-black Necker cube projected onto a black screen. The total number of apparent shifts for two 110-second sessions was used as the *S's* score.

b. *Spiral visual aftereffect (SAE)*. A 7-inch-diameter spiral of Archimedes with a luminous paint surface and illuminated by an ultraviolet light was used as the test stimulus. The spiral was rotated by a constant-speed 100-RPM Bodine motor for a 30-second period. *S* indicated the duration of the SAE by pressing a signal button while the aftereffect lasted. The sum of the SAEs in the two sessions, each SAE measured to the closest one-tenth second, was used as the *S's* score.

c. *Phi phenomenon*. Two electroluminescent one-half inch diameter circular panels separated 30 inches were used as stimuli. Lights were activated alternately by a micropositioner with the frequency determined by a Hewlett-Packard Ultra-Low-Frequency Generator. By use of the psychophysical method of limits, a threshold frequency (to the closest one-tenth cycle) was determined for the point at which the phi phenomenon just began to appear or disappear. The sum of the two threshold frequencies was used as the *S's* score.

d. *Autokinetic effect*. A pinpoint low-intensity light source was used as the stimulus. *S* observed the light for a 110-second period and indicated the apparent duration of movement by pressing the signal button for as long as the effect continued and for each time it appeared. The sum of the two session durations was used as the *S's* score.

### 3. Procedure

The *S* was first instructed concerning the nature of visual illusions, and these effects were demonstrated. All the *Ss* were then tested for each illusion for two sessions under a neutral-set-and-no-shock condition. The sum of the two session scores represented the base-line (Condition 1) measure for each *S*. Its effect was removed from the experimental effects (Condition 2) by analysis of covariance.

Following a 20-minute rest, *Ss* were then tested with these same illusions



under experimental conditions (Condition 2). Seven of the Ss were randomly assigned to each of nine experimental conditions: one of three shock levels (none, mild [three milliamps for three-second duration], and strong [five milliamps for five-second duration]) combined with one of three perceptual sets (facilitating the perception of the illusions, neutral set, and inhibiting the perception of the illusions). Electric shock was presented immediately before each trial for each illusion.

### C. RESULTS AND DISCUSSION

A reliability estimate for each of the four illusions was determined by correlating the Ss' two session scores of Condition 1. Coefficients of .84 for the Necker cube, .86 for the SAE, .63 for the phi phenomenon, and .83 for the autokinetic effect are evidence of high reliability in the measurement of such perceptual performance.

When performance under Condition 1 for each illusion was compared with that for each of the other illusions, insignificant correlations were obtained. Such a result indicated that exposure to one perceptual task had no consistent bias in modifying performance for the other tasks.

For each of the four illusions, an analysis of variance for the Condition-1 scores of the Ss assigned to the nine experimental conditions revealed no significant group differences. This result is evidence that significant effects obtained under Condition 2 are due to the experimental conditions rather than due to initial individual differences among Ss in each group.

Table 1 summarizes the *F* ratios based on an analysis of covariance of shock-set effects with Condition-1 effects removed. It is seen that mental set significantly influences perceptual performance for the Necker cube, SAE, and autokinetic effect, while shock stress significantly influences perceptual performance of the autokinetic effect.

It seems unlikely that the insignificant shock effect observed for the Necker cube, SAE, and phi phenomenon is due to the relatively low shock levels used.

TABLE 1  
ANALYSIS OF COVARIANCE OF SET AND SHOCK EFFECTS ON ILLUSION PERCEPTION  
FOR FOUR VISUAL ILLUSIONS

Source	df	F			
		Necker cube	SAE	Autokinetic effect	Phi phenomenon
Set	2	15.20**	10.00**	3.74*	1.62
Shock	2	1.50	0.33	3.27*	2.02
Set $\times$ Shock	4	0.46	0.38	0.13	1.24
Within	53				
Total	61				

\*  $p < .05$ .

\*\*  $p < .01$ .



After completing the experiment, each *S* rated the shock intensity on a five-point scale. Those *Ss* who had received a strong shock consistently rated the shock as "very strong." The verbal distress observed by *E* over the intercom also seemed to support such ratings. However, individual differences in reaction were more apparent for the mild-shock stress condition.

Shaffer (11), in his "neurophysiological hypothesis" concerning stress, proposed that stress should functionally decorticate the organism in such a manner that there is a breakdown of communication between the cortex and the lower centers of the brain. In the present experiment, mental set may have inhibited such disorganization in the cognitive processes due to the stress effects. The fact that shock stress had little effect for "normal" *Ss* in no way contradicts the finding that severe shock produces a decrement in perceptual performance when applied to psychotics (3).

The mean performances for Condition 2 for the nine experimental groups are presented in Figure 1. It is seen that, when compared to a neutral set and for all shock levels, a positive set results in greater perceptual performance, while a negative set results in less perceptual performance. Maximum set differences occur under strong-shock conditions, while minimum set differences occur under mild-shock conditions for the Necker cube, SAE, and autokinetic effect.

For the Necker cube, contrary to the predicted outcome when compared to a no-shock condition, strong shock increases—while mild shock decreases—perceptual performance. Perhaps strong shock may alert the *S* to the task at hand, while mild shock may merely distract and not arouse the *S*.

For the phi phenomenon, mild shock increases—while strong shock decreases—threshold scores, as measured in cycles per second. However, since a higher threshold represents less illusory effect, the observed outcome (in terms of the constricting effects of stress) is the same as that obtained for the Necker cube and the SAE. Since strong shock resulted in less difference in the three set groups, it would seem that shock should be an important determiner of such perceptual performance. Perhaps the insignificant set and stress effects were due in part to the lower reliability of the measures, as well as to the experimental method of determining phi thresholds. Although the psychophysical method of constant stimuli would have been more reliable than the method of limits, the time requirements of such a procedure would have increased the duration of the experiment beyond practical limits.

Of the four illusions, the most atypical outcome was that obtained for the autokinetic effect. Both set and shock effects were significant, with mild stress increasing the duration of the effect for all three sets, and strong shock

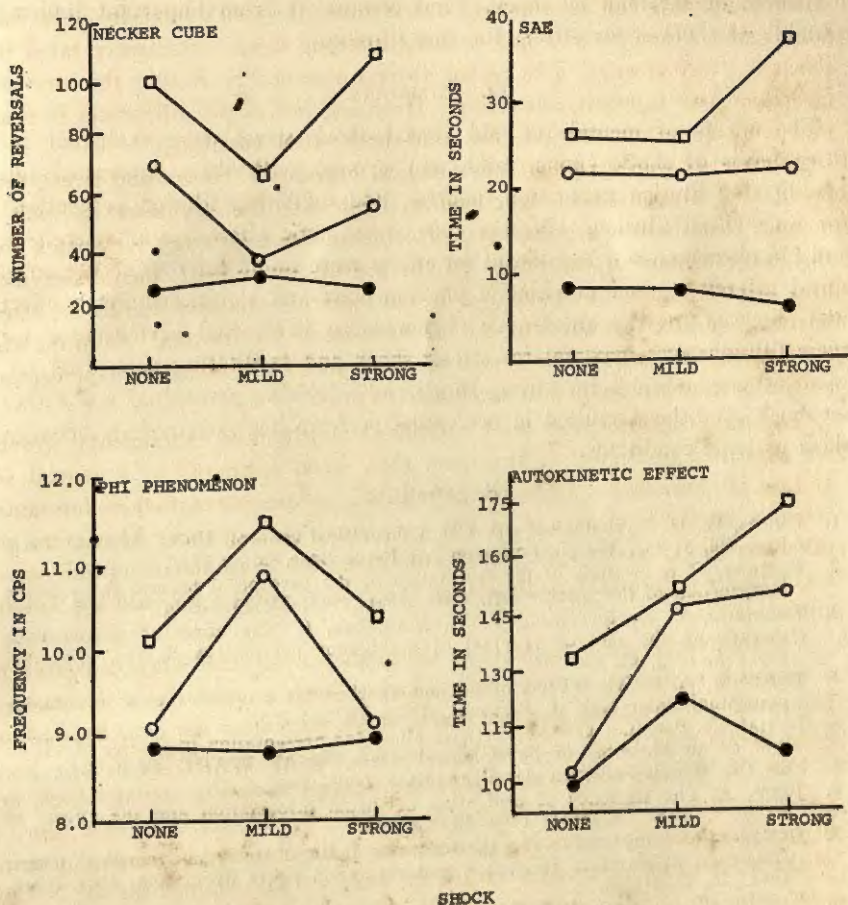


FIGURE 1

MEAN SCORES FOR THREE SHOCK LEVELS FOR THREE SET CONDITIONS FOR FOUR ILLUSIONS  
 Positive mental set condition is represented by squares, neutral set by circles, and negative mental set by closed circles.

further increasing the duration for positive and neutral sets and decreasing the duration for the negative set. Such an outcome is evidence that the auto-kinetic effect may be a useful stress-sensitive measure in future research. If stress is essentially a disrupting experience, then the longer duration of auto-kinetic effect under strong shock conditions represents greater stress effects. However, since strong-shock negative-set conditions result in decreased duration of the illusory effect, then perception under such conditions is more



influenced by set than by shock. Thus mental set is an important limiting variable of shock-stress effects for this illusion.

#### D. SUMMARY

The effects of mental set and electric-shock stress were evaluated for three levels of shock (none, mild, and strong) with three perceptual sets (facilitating illusion perception, neutral, and inhibiting illusion perception) for four visual illusions (Necker cube, autokinetic effect, spiral aftereffect, and phi phenomenon). Significant set effects were found for the Necker cube, spiral aftereffect, and autokinetic phenomenon; and significant shock effect was observed for the autokinetic phenomenon. Perceptual performance for these illusions was maximal for strong shock and facilitating perceptual set, while it was minimal for strong shock and inhibiting perceptual set. Other set-shock conditions resulted in perceptual performance intermediate between these extreme conditions.

#### REFERENCES

1. CHILES, W. D. Psychological stress as a theoretical concept. USAF WADC Tech. Rep. No. 57-457, Wright-Patterson Air Force Base, Ohio, 1957.
2. COSTILOE, C. S. A study of the facilitation of the perceptual process through the introduction of the specific sets. *Diss. Abst.*, 1955, 15, 2306-2307.
3. FISICHELLI, V. R., ROCKWELL, F. V., & CLARKE, L. The effect of electro-shock therapy on the rate of reversal of ambiguous perspective figures. *Amer. J. Psychol.*, 1955, 68, 638-644.
4. HEILBRUN, A. B. The effects of various shock-stress conditions upon a complex perceptual-motor task. *J. Personal.*, 1959, 27, 285-299.
5. HYMAN, A. Potential uses of alternate binocular presentation in studies of vision and as an indicator of physiological stress. USAF WADC Tech. Rep. No. 60-302, Wright-Patterson Air Force Base, Ohio, 1960.
6. JONES, A. The efficiency of utilization of visual information and the effects of stress. *J. Exper. Psychol.*, 1959, 58, 428-432.
7. KUSHNER, M. The effects of experimentally induced stress upon critical flicker frequency, palmer skin resistance, and reversible figure fluctuation. *Diss. Abst.*, 1955, 15, 2583-2584.
8. LAZARUS, R. S., DEESZ, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, 49, 293-316.
9. LONG, E. R., HENNEMAN, R. H., & REID, L. S. Theoretical considerations and explanatory investigation of "set" as a response restriction. USAF WADC Tech. Rep. No. 53-311, Wright-Patterson Air Force Base, Ohio, 1953.
10. SALTZ, E., & RIACH, W. J. The effect of stress on stimulus differentiation. *J. Exper. Psychol.*, 1961, 62, 588-593.
11. SCHAFFER, H. R. Behavior under stress: A neurophysiological hypothesis. *Psychol. Rev.*, 1954, 61, 323-333.
12. WEINER, M. G. Perception under mild stress: Factors involved in changes in recognition threshold under electric shock. *Clark Univ. Bull.*, 1949, No. 192, 83-86.

Department of Psychology  
Boston College  
Chestnut Hill 67, Massachusetts





